ESSAYS ON STRUCTURAL ANALYSIS OF PROCUREMENT AUCTIONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Bin Yu, M.A.

* * * * *

The Ohio State University

2007

Dissertation Committee:
Professor Lung-fei Lee, Adviser
Professor Stephen R. Coslett
Professor Lixin Ye

Approved by

[Signature]
Adviser
Economics Graduate Program
ABSTRACT

This dissertation addresses the empirical analysis of procurements based on the auction theory, which is known as the structural-form analysis of procurement auctions.

The first essay studies the nonparametric estimation of asymmetric procurement auctions and the empirically implementable bandwidth selection methods in the finite-sample estimation of procurement auctions with heterogeneity in the project sizes. Guerre, Perrigne and Vuong (2000) (GPV) proposed a two-step nonparametric estimator and established its property of uniform convergence under the appropriate bandwidth choices. However, their theoretical bandwidth falls short of providing the practical guidance for empirical applications. In this essay we develop an empirical bandwidth selection rule for the nonparametric estimator under the GPV specification. We also propose a conditional hazard rate based nonparametric estimator and develop its empirical bandwidth selection rule. By conducting Monte Carlo experiments of a simulated auction model, we demonstrate that when the heterogeneity of the project size exists as in any typical auction data, the proposed bandwidth selection rules under both the GPV and the conditional-hazard-rate based specifications produce on average much improved estimates of the unobserved cost and its distribution than the commonly adopted bandwidth selection rule in the literature. Further, the conditional hazard rate based specification is more preferred in this context. Then we
apply the conditional hazard rate based nonparametric estimator with its bandwidth selection rule to an empirical analysis of the Ohio highway procurement market. Two types of firms, regular or fringe according to the frequency of bidding, are studied in terms of their respective markups and the project-dependent cost information structures in this market. In contrast to the usual point of view that regular firms have a cost advantage over fringe firms, our finding is that this conclusion holds only when the project is of relatively large size. In projects of small size, fringe firms can enjoy a cost advantage.

The second essay studies the identification and estimation of procurement auction models with the endogenous entry of potential firms. Two different procurement auction models with endogenous entry are studied. The procurement auction model with the bid-preparation cost follows from an irrevocable and nontrivial entry expense, which is incurred after a potential firm observes its private project cost. The procurement auction model with the information-gathering cost follows from an irrevocable and nontrivial entry expense, which is incurred before a potential firm observes its private project cost. In this essay we show how each of these two auction models with the endogenous entry can be identified and estimated from the data. We also establish the identification result in a broader sense: we can differentiate these two models by only using the data of observed bids. Finally, we empirically analyze the Michigan highway procurement market where the endogenous participation of potential firms can be justified. By applying the established identification criterion and checking the model implications, we conclude that the procurement auction model with the bid-preparation cost is more appropriate to explain the observed participation and
bidding strategies of the potential firms in the Michigan highway procurement market. We also use the Bayesian method to recover the project cost distribution and the entry cost distribution in this market.
Dedicated to my parents and my wife.
ACKNOWLEDGMENTS

First, I would like to express my gratitude to my adviser, Professor Lung-fei Lee. The knowledge I learned in his class and the ideas inspired during his insightful talks made me interested in research in econometrics. Throughout my study in the past several years, he gave me constant guidance, invaluable advice and warm encouragement. It is his intellectual support that makes this thesis possible. He also showed valuable patience and prudence in correcting the mistakes in my paper. His admirable personality and professional ability are of great benefit to my study and my life in the future.

Second, I want to thank my dissertation committee members: Professor Steven R. Cosslett and Professor Lixin Ye. Their valuable suggestions and comments really helped me improve this dissertation. My thanks also go to Professor Joseph P. Kaboski. The experience of being his research assistant helped me gain understanding in research outside my field and extend my mind. The computational skills I learned during work with him smoothed my way of solving the problems in my dissertation afterwards.

And I appreciate all the intellectual advices and responsible directions from Hajime Miyazaki. I am also grateful to Jo Ducey, Ana Shook and John-David Slaughter for their assistance all the time.
My special thanks are dedicated my father Daming Yu and my mother Yunmei Huang. Their trustful love supported me in every stage of my life. Without them, none of my achievements would have been possible. I also want to thank my wife, Xiaojing Li. Her unconditional support and deepest love makes me a happier person. She is the best blessing in my life.
VITA

November 27, 1977 .......................... Born - Yidu, China

July, 2000 ................................. B.A., Economics, Wuhan University, China
July, 2000 ................................. B.S., Mathematics, Wuhan University, China
August, 2003 ............................. M.A., Economics, The Ohio State University

FIELDS OF STUDY

Major Field: Economics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Vita</td>
<td>viii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Parametric vs. Nonparametric Methods</td>
<td>8</td>
</tr>
<tr>
<td>1.2 Exogenous vs. Endogenous Participation</td>
<td>11</td>
</tr>
<tr>
<td>2. PRACTICAL BANDWIDTH SELECTION METHOD IN NONPARAMETRIC ESTIMATION OF PROCUREMENT AUCTION MODEL</td>
<td>14</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>14</td>
</tr>
<tr>
<td>2.2 The Procurement Auction Model with Heterogeneous Firms</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Bandwidth Selection under GPV Nonparametric Specification</td>
<td>20</td>
</tr>
<tr>
<td>2.4 Bandwidth Selection Under Conditional Hazard Rate Based Nonparametric Specification</td>
<td>25</td>
</tr>
<tr>
<td>2.5 Comparison between Different Methods and Monte Carlo Experiment</td>
<td>29</td>
</tr>
<tr>
<td>2.6 Application to the Ohio Highway Procurement Market</td>
<td>35</td>
</tr>
<tr>
<td>2.6.1 Description of the Ohio Highway Procurement Market</td>
<td>36</td>
</tr>
<tr>
<td>2.6.2 The Data and Some Reduced-Form Analysis</td>
<td>37</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Bandwidth Used in Estimating Inverse Bidding Functions</td>
<td>106</td>
</tr>
<tr>
<td>Table 2</td>
<td>CMISE/MSE of Estimated Inverse Bidding Functions</td>
<td>106</td>
</tr>
<tr>
<td>Table 3</td>
<td>Explanation of Variables in Ohio Highway Procurement Market</td>
<td>107</td>
</tr>
<tr>
<td>Table 4</td>
<td>Summary Statistics of Main Variables</td>
<td>107</td>
</tr>
<tr>
<td>Table 5</td>
<td>Division of Firms by the Number of Total Bids</td>
<td>107</td>
</tr>
<tr>
<td>Table 6</td>
<td>Division of Projects by Size</td>
<td>108</td>
</tr>
<tr>
<td>Table 7</td>
<td>OLS and Probit Regression Results</td>
<td>108</td>
</tr>
<tr>
<td>Table 8</td>
<td>Estimated Markups in Different Subgroups</td>
<td>108</td>
</tr>
<tr>
<td>Table 9</td>
<td>Summary Statistics of the Normalized Cost</td>
<td>110</td>
</tr>
<tr>
<td>Table 10</td>
<td>OLS Regression for Recovered Cost</td>
<td>110</td>
</tr>
<tr>
<td>Table 11</td>
<td>Summary Statistics for Michigan Highway Procurement Market</td>
<td>110</td>
</tr>
<tr>
<td>Table 12</td>
<td>Regression Analysis on Number of Actual Bidders</td>
<td>111</td>
</tr>
<tr>
<td>Table 13</td>
<td>Regression Analysis on Number of Potential Bidders</td>
<td>111</td>
</tr>
<tr>
<td>Table 14</td>
<td>Summary of the Effects of N/n on LogBid and Nbids</td>
<td>112</td>
</tr>
<tr>
<td>Table 15</td>
<td>Regression for the Dependent Variable: Actual Number of Bidders</td>
<td>112</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1: Estimates of the Inverse Bidding Functions</td>
<td>114</td>
</tr>
<tr>
<td>Figure 2: Estimate of the Conditional Density Function of Private Cost</td>
<td>116</td>
</tr>
<tr>
<td>Figure 3: Geographical Distribution of Projects and Firms by County</td>
<td>116</td>
</tr>
<tr>
<td>Figure 4: Plot of the Normalized Bids</td>
<td>117</td>
</tr>
<tr>
<td>Figure 5: Plot of Bid vs. Eng.Est.</td>
<td>117</td>
</tr>
<tr>
<td>Figure 6: Histogram of the Normalized Cost Estimates</td>
<td>118</td>
</tr>
<tr>
<td>Figure 7: Density Function of the Normalized Cost Estimates</td>
<td>119</td>
</tr>
<tr>
<td>Figure 8: An Illustrative Example for Model with Bid-preparation Cost</td>
<td>120</td>
</tr>
<tr>
<td>Figure 9: An Illustrative Example for Model with Information-gathering Cost</td>
<td>120</td>
</tr>
<tr>
<td>Figure 10: Histograms of Selected Explanatory Variables in Michigan Market</td>
<td>121</td>
</tr>
<tr>
<td>Figure 11: Distribution of the Regression Errors</td>
<td>121</td>
</tr>
<tr>
<td>Figure 12: Time Series Plot of Latent Variables for a Procurement Auction</td>
<td>121</td>
</tr>
<tr>
<td>Figure 13: Bayesian MCMC Simulation Results</td>
<td>121</td>
</tr>
<tr>
<td>Figure 14: Comparison between Observed and Simulated Procurement Costs and Number of Bidders</td>
<td>123</td>
</tr>
<tr>
<td>Figure 15: Effect of the Number of Potential Bidders on Bid</td>
<td>123</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

As a simple and well-defined economic institution, auction plays an important role in the worldwide economy. Auctions are widely used to sell goods or services. Examples include antiques or artworks auctions with long history, as well as government-sponsored auctions to allocate natural resources such as timber and oil. Auctions have become even more popular and shaped new consumer shopping patterns since the burgeoning of internet auctions. With auctioned objects ranging from baby feeding supplies to cars and real estate, Ebay had total sales of $44.3 billion from the 1.9 billion listings in 2005 and a global cumulative customer base of 180.6 million at the end of 2005, making it the biggest online auction marketplace. On the other hand, another economically important auction mechanism is procurement auction, which is used to purchase goods and services. Bidders in procurement auctions are usually businesses or cooperations. For example, many state governments in the U.S. use procurement auctions to contract out highway construction projects and purchase office supplies. In 2002, the total procurement expenditures by the U.S. federal government were $271 billion, which accounted for 14% of all government expenditures in the same year.
The wide application of auction mechanism has led to advanced theoretical study of auction, which provides us with a setting to analyze the economic agents’ non-cooperative strategic interactions under incomplete information and helps us to understand more complicated economic institutions. Auction theory makes use of game theory to rationalize bidders’ strategies and study the design of optimal auction mechanisms. More specifically, the procurement auction\(^1\) has been modeled as a non-cooperative game under incomplete information\(^2\). In the low-price sealed-bid independent private cost (IPC) paradigm, which has been most thoroughly studied in theory and widely applied in practice, \(n\) bidders compete for an indivisible project and the bidder with the lowest bid wins the contract. Each bidder observes its own project cost and does not know the project costs of other bidders. However, all bidders’ project costs are modeled as independent random realizations from a distribution \(F(\cdot)\) on the support \([c, \bar{c}]\). Knowing the private project cost \(c\), the number of bidders \(n\), the distribution of project cost \(F(\cdot)\), each bidder submits a bid \(b\) to maximize the expected profit \(E(\Pi_{\text{bidder}})\), written as

\[
E(\Pi_{\text{bidder}}) = (b - c)F^{-1}(s^{-1}(b))
\]

where \(s(\cdot) = s(\cdot|n, F)\) is the assumed Bayesian Nash Equilibrium (BNE) bidding strategy and in general depends on the number of bidders \(n\) and the distribution of project cost \(F(\cdot)\). Given the number of bidders \(n\) and the equilibrium bidding

\(^1\)Our analysis will be focused on the procurement auction, which is also known as low-price auction since the winner is the bidder with the lowest bid. The analysis can be extended to the first-price auction, in which the winner is the bidder with the highest bid.

function \( s(\cdot | n, F) \), which can be explicitly solved as

\[
s(c|n, F) = c + \frac{\int_c^\pi [1 - F(t)]^{n-1} dt}{[1 - F(c)]^{n-1}},
\]

the procurer’s expected contract cost conditional on the number of bidders \( E(C_{\text{procurer}} | n) \) from the low-price sealed-bid auction mechanism is

\[
E(C_{\text{procurer}} | n) = \int_\xi s(x|n, F)dF^{(1)}(x)
\]

where \( F^{(1)}(x) = 1 - [1 - F(x)]^n \) is the distribution of the lowest order statistics of \( n \) random variables identically and independently drawn from the distribution \( F(\cdot) \). If the procurer chooses a different auction mechanism, the bidder’s equilibrium bidding strategy and procurer’s expected contract cost conditional on the equilibrium strategy will change accordingly. From the perspective of a bidder, the theory literature studies how the bidder chooses the optimal bid to maximize the expected profit. From the perspective of the procurer, the theory literature studies how different auction mechanisms affect the procurer’s expected contract cost. The theory literature also studies how different auction mechanisms can affect the total social surplus from the perspective of the central planner. The latter two studies are related to the design of optimal auction mechanism.

The well established auction theory and widely available auction data have resulted in extensive empirical study of auctions. Two main approaches exist in the empirical auction literature: the reduced-form analysis and the structural-form analysis. The reduced-form analysis has a long history and is intuitive in the sense that it studies the relationship between the response variable (i.e., the observed bid \( b \)) and
the project-specific and/or bidder-specific covariate vector $X$ and the number of bidders $n$. This approach usually involves a linear regression model and the randomness is imposed from outside the model. More specifically, the reduced-form analysis can be represented by one of the following regression equations:

$$b = X\beta + \alpha \cdot n + \epsilon$$  \hspace{1cm} (1.1)

$$b = X\beta + \alpha \cdot \ln(n) + \epsilon$$

The reduced-form analysis can be used to test assumptions and implications of the theoretical auction model, explain the variations of bids between auctions and predict bids in future procurement auctions. For example, the regression model above can be used to test whether the observed bid decreases when the number of bidders increases in the low-price sealed-bid IPC model.

The structural-form analysis started from Paarsch (1990, 1992) and has been a very active research area in the fields of industrial organization and econometrics since then. Different from the reduced-form analysis, the structural-form analysis studies the relationship between the latent variable (i.e., the unobserved project cost $c$) and the project-specific and/or bidder-specific covariate vector $X$ and the distribution of the latent variable. This approach is developed from the theoretical auction model which well defines a self-contained structure with randomness. By assuming that the observed bids come from the BNE bidding strategy, an econometric model can be developed to estimate the structural elements in the model such as the distribution function $F(\cdot)^3$. For example, to decide between common and private value paradigms for the tree planting contract auctions, Paarsch estimated two simple theoretical paradigms.

\footnote{For more comprehensive survey on the existing literature in the structural analysis of auction models, please refer to Perrigne and Vuong (1999), Sareen (2002).}

4
models of rational bidding in sealed-bid auctions. In the common value model, the contract is assumed to have a common but unknown cost \( c \) to all bidders ex ante. Each bidder only observes a private estimate \( x \) for the unknown cost \( c \), which is assumed to be drawn independently from a distribution function \( F(\cdot|c) \). The prior distribution \( G(\cdot) \) for \( c \) is assumed to be known to all bidders. Two parametric assumptions for the underlying model \([F(\cdot|c), G(\cdot)]\) are considered: the Rothkopf-Smiley parametric specification which guarantees the bid \( b \) is proportional to the estimate \( x \) and Levin-Smith parametric specification which assumes the unbiased normal estimate, i.e., \( F(\cdot|c) \) is a normal distribution with mean \( c \), and the priors for the unknown value \( c \) are identical and diffuse. In the private value setting, the contract is assumed to have a different private cost to each bidder. Three parametric assumptions for the underlying model are considered that respectively guarantee that bid is proportional to private cost, the bid is additive in private cost and the bid is nonlinear in private cost. After fitting the empirical models under different parametric assumptions by estimating the relevant parameters, the author concluded that the common value paradigm is more applicable for the data from tree planting contract auctions, with Rothkopf-Smiley specification more preferred. This empirical conclusion from the structural analysis is of great practical importance. According to the results of Milgram and Weber (1982), the procurer will incur the highest expected contract cost from low-price sealed-bid auction under common value paradigm. So an English auction or a second low price auction should be used instead to procure the contract.

In the structural-form analysis, researchers are interested in studying how the unobserved project cost depend on the project-specific and/or bidder-specific covariates and how it is distributed. So a parametric structural econometric auction model to
be estimated can be specified by a two-equation model:

\[ c^* \sim F(\cdot|X, \beta) \]

\[ b = s(c^*|n, F) \]

where \( s(\cdot|n, F) \) is the bidding strategy adopted by bidders. In this two-equation specification, we have implicitly assumed that the unobserved cost \( c^* \) doesn’t depend on the number of bidders \( n \) in an auction; but as derived from the BNE bidding strategy, the bidding function \( s(\cdot|n, F) \) depends on the number of bidders \( n \) and the distribution function \( F(\cdot|X, \beta) \). In the mathematical terms, the structural auction model can be described as a setting in which the observed response variable \( b \) is determined by the unknown latent variable \( c^* \) through a possibly highly nonlinear function, which depends on the unknown distribution function \( F(\cdot|X, \beta) \) from which the latent variables are drawn and the number of bidders \( n \). By estimating the parameter vector \( \beta \) and the distribution of \( c^* \), we can not only predict bid as in the reduced-form analysis, but also recover the unobserved cost \( c^* \) corresponding to each \( b \) and the distribution function of \( c^* \), which can then be used to evaluate the efficiency of the current auction mechanism or design better auction mechanism.

Without rationalizing the bidders’ strategies and relating the observed response variable with the unobserved latent variable, the reduced-form approach usually focuses on a specific auction setting and cannot identify the latent project cost generating mechanism implied by auction theory. Thus it cannot make inference outside of the specific auction setting. On the contrary, the structural-form analysis identifies the latent project cost distribution, which is the basis of any auction mechanism. Thus it can be used to make inference in other auction mechanisms and compare alternative auction mechanisms. The structural-form analysis of auction models is
closely related with the optimal auction design. The feasibility of the optimal auction
design in practice relies on the assumption that researchers or policy makers have
enough information about the distribution of the project cost. For example, in the
IPC paradigm as described above, if the procurer wants to set a reservation price
to minimize the expected contract cost, the optimal reservation price will depend
on the distribution function \( F(\cdot) \). However, the information about the distribution
function \( F(\cdot) \) is usually not available to researchers or policy makers. Solving this
lack-of-information problem for evaluating or designing an auction mechanism has
been one of the main objectives in developing structural analysis of auction models.
The structural analysis of auction models is an approach to integrate theory with
practice and "put auction theory to work". Econometric auction models are devel-
oped in structural analysis to obtain the information about the distribution function
\( F(\cdot) \) from past procurement practices, which can then be used to evaluate the current
auction mechanism or to design other auction mechanisms with considerations to cost
minimization, efficiency and/or fairness. The following proposition establishes the es-
sential difference between the reduced-form and structural-form approaches from an
econometric perspective in the symmetric IPC procurement auction setting.

**Proposition 1** In the symmetric IPC procurement auction model, suppose the project
cost \( c^* = X\beta + (Z\gamma) \cdot e \) where \( X \) and \( Z \) are the covariate vectors to characterize the
observed heterogeneity of different projects and \( e \) is the standardized error term with
mean 0 and variance 1 that characterizes the unobserved component of project costs
of different firms, which is independent of both the covariate vectors and the number
of bidders and is distributed according to \( F(\cdot) \) on the support \([c, \bar{c}]\). Then for the
equilibrium bid \( b = s(c^*|X, Z, n, F) \), we have

\[
E[b|X, Z, n] = X\beta + (Z\gamma) \cdot g(n)
\]

where \( g(n) \equiv E[s_0(e|n, F)] \) and \( s_0(\cdot|n, F) \) is the BNE bidding strategy given \( X\beta = 0, Z\gamma = 1, n \) and \( F(\cdot) \). Furthermore, there exists no distribution function \( F(\cdot) \) such that \( g(n) = a + b \cdot n \) or \( g(n) = a + b \cdot \ln(n) \) for some constants \( a \) and \( b \).

The proposition above shows that in the symmetric IPC procurement auction, given the parametric specification and the distributional assumption for \( c^* \), we can consistently estimate the unknown parameter \( \beta \) and \( \gamma \) and thus the project cost distribution by using a OLS regression. GLS/FGLS can also be used to improve the efficiency of the estimates. The proposition also shows the reduced-form estimation as specified by (1.1), where the number of bidders enter the regression equation linearly or log-linearly, can never be rationalized in the structural-form estimation.

1.1 Parametric vs. Nonparametric Methods

The structural econometric auction models can be estimated by both parametric and nonparametric methods. Although the relationship between the response variable \( b \) and the latent variable \( c^* \) in many auction settings is easy to be characterized as shown by the auction theory literature, estimation of the distribution of project cost is nontrivial. In a parametric setting, since the BNE bidding function in general depends on the parameters in the distribution of project cost, the support of the response variable \( b \) will depend on those parameters as well. This makes the assumptions for establishing the consistency and asymptotic normality of the standard

\(^4s_0(\cdot|n, F) \) is thus a normalized bidding function in the mathematical sense since the project cost cannot be negative in practice.
maximum likelihood estimator break down. Many alternative methods have been proposed. For example, Donald S.G. and Paarsch H.J. (1993) proposed the piecewise pseudo-maximum likelihood estimation method. Donald and Paarsch (1996) further proposed the constrained maximum likelihood estimator. Laffont, Ossard and Vuong (1995) proposed simulated nonlinear least square (SNLLS) estimator for the first-price sealed-bid IPV auction using the winning bid. Li and Vuong (1997) proposed a similar SNLLS estimator using all bids. Li (2005) proposed a simulated GMM method to estimate a parametric first-price two-stage auction model with entry cost and binding reservation price. However, all these methods have made use of either an explicitly expressed bidding function or the revenue equivalence theorem in auction theory in the estimation. If we extend the auction model by considering asymmetric types of bidders, it becomes difficult to carry through these methods because the BNE bidding functions cannot be explicitly expressed\(^5\) and the revenue equivalence theorem does not hold any more.

In terms of computational complexity, the nonparametric method will have much more advantage over the parametric method. Since the leading work by Guerre et al. (2000), the nonparametric method has been widely used in the structural estimation of the auction model. It makes use of the observed data such as the bidder-specific bids and the auction-specific covariates to recover the distributions of project cost without assuming any specific form of the underlying distributions. In addition to the robustness to misspecification, the nonparametric method is easy to implement since it does not solve any differential equation or numerical integration and thus

\(^5\)The bidding functions of different types of bidders are the solutions to a system of first-order differential equations with appropriate boundary conditions and usually cannot be expressed explicitly.
involves much less numerical burden; the nonparametric method also can be easily extended to the auction model with asymmetric types of bidders. The nonparametric identification and estimation in the structural auction model result from the fact that although the bidding function itself is unknown, its inverse function can be easily obtained, which explicitly expresses the unknown project cost as a function of the observed bid. The form of this inverse bidding function depends on the number of bidders and the bid distribution function which can be estimated from the data. Following Guerre et al. (2000), the nonparametric identification and estimation in the structural form have been extended to many other auction models. Li, Perrigne and Vuong (2000) studied the conditionally independent private value auction model. Krasnokutskaya (2002) studied the highway procurement auction under unobserved heterogeneity. Campo, Perrigne and Vuong (2003) studied the first-price auction with affiliated private values with asymmetric types of bidders. Li & Perrigne (2003) studied the auction with random reservation price. Flambard and Perrigne (2004) studied procurement auctions with asymmetric types of bidders.

Although the nonparametric method is straightforward and easy to implement in terms of numerical complexity, we are confronted with a crucial step to assess the bidders’ unobserved project costs accurately in applying kernel nonparametric estimation method in the empirical analysis: choose an appropriate bandwidth. This is important in the structural auction model since the empirical conclusions and policy analysis are based on the nonparametric estimates. Theoretical bandwidth choices proposed by Guerre et al. (2000) in establishing the optimal convergence rate fall short of providing a practical guidance for the empirical applications since they consist of undetermined constants. Although the nonparametric method has been widely
used in the empirical auction analysis, the literature has paid little attention to the bandwidth selection in finite sample estimation, especially when there exists heterogeneity in the procured projects as in any typical auction data. In chapter 2 we propose practical bandwidth selection rules in the two-step nonparametric estimation of procurement auction model. When there exists heterogeneity across different projects, the proposed bandwidths improve estimation results as compared to the commonly adopted bandwidths in the literature. We compare different estimation methods by using a Monte Carlo experiment of simulated procurement auctions. We also apply the proposed bandwidth selection methods to an empirical analysis of the Ohio highway procurement auctions.

1.2 Exogenous vs. Endogenous Participation

In most structural form analysis of auction models, researchers do not consider potential bidders’ participation strategy and thus do not differentiate between potential bidder and actual bidder. As in the low-price sealed-bid IPC model discussed above, it is usually assumed in the empirical analysis that the number of actual bidders is a common knowledge among bidders and does not depend on the distribution of the project cost. These assumptions have made the analysis of bidding strategy easier and the estimation of the latent variable distribution more tractable. However, in the empirical analysis, the assumption that the number of bidders is publicly known is questionable, especially in the sealed-bid auction. A more realistic model need to differentiate a potential bidder from an actual bidder and consider potential bidders’ participation strategies. The most obvious reason for the endogenous participation is the existence of an effective reservation price. In this situation, only
those potential bidders who have project costs no larger than the reservation price will actually submit their bids. In other words, the bidding firms’ project costs are truncated below the reservation price. It follows that the number of bidding firms will be endogenously determined. Guerre et al. (2000) established the nonparametric identification result and estimation method for the first-price sealed-bid auction model under the assumptions that the number of potential bidders is constant and the reservation price is a possibly unknown deterministic function of the observed auction characteristics. Flambard and Perrigne (2005) analyzed the snow removal contracts where the endogenous entry resulted from an effective reservation price. Given the facts that the reservation price is usually observable and the BNE participation and bidding strategies in this model are straightforward, the identification and estimation of the structural auction model is not difficult. The more interesting and complicated auction model with endogenous participation results from the existence of entry cost. In chapter 3 we will study structural inference on procurement auction models with entry cost. Two auction models with endogenous entry are considered. In the first case, the endogenous entry results from the bid-preparation cost, which is incurred after the bidder observes its project cost. In the second case, the endogenous participation results from the information-gathering cost, which is incurred before the bidder observes its project cost. Since the entry cost is incurred at different stages of the procurement auction process, the bidders in these two models have different participation and bidding strategies and these two models impose different restrictions on the observed data. In this chapter, we will study the identification of these two auction models with entry both separately and jointly. We also empirically study
the Michigan highway procurement auction market by using the procurement auction model with the bid-preparation cost.

The dissertation is organized as follows. Chapter 2 studies bandwidths selection methods in the two-step nonparametric estimation of the structural procurement auction model under finite sample from heterogeneous auctions. We apply the proposed bandwidth selection rules to an empirical analysis of Ohio highway procurement auctions. Chapter 3 considers the structural analysis of procurement auction models with entry. We also empirically analyze the Michigan highway procurement auctions. The proof of propositions, explanation of Bayesian MCMC estimation method used in chapter 3, tables and figures are included in the appendix.
CHAPTER 2

PRACTICAL BANDWIDTH SELECTION METHOD IN NONPARAMETRIC ESTIMATION OF PROCUREMENT AUCTION MODEL

2.1 Introduction

Since the pioneering work by Guerre et al. (2000)\(^6\), the nonparametric method has been widely used in the structural-form estimation of auction models. In the procurement auction setting, the observed data, which includes the firm-specific bids and the project-specific covariates, are used to estimate the firms’ unobserved project costs without assuming any parametric form for the underlying project cost distribution. Compared to the parametric method, the major advantages of the nonparametric method are that it is robust to misspecification and easy to implement since it does not involve solving differential equations or calculating complicated integrations. However, the nonparametric method requires the researchers to select an optimal bandwidth, and more importantly, a practical bandwidth when the asymptotic theory does not provide enough guidance in the finite sample estimation. The bandwidth selection in the finite sample estimation is critical for assessing firms’ project costs accurately for the policy-orientated analysis. The practical issue that

\(^6\)We use GPV as a simplification hereafter.
makes the bandwidth selection more involved is that in contrast to the experimental auction data which can be well controlled, the field auction data are usually characterized by heterogeneity in several dimensions. First, the number of bidders varies from one auction to another. Second, the project size usually differs in any two different contracts. The commonly adopted bandwidth in the empirical auction literature is not robust to the heterogeneity of project size in the finite sample estimation. In this chapter we develop practical bandwidth selection rules when the underlying project cost distribution is conditioned on a variable that characterizes the heterogeneity of project size. This chapter has connections with two fields: the nonparametric estimation of the conditional hazard rate function and the conditional density function and the structural-form estimation of auction models. The organization of this chapter is as follows: section 2 presents the basic procurement auction models with two asymmetric types of firms; in section 3, we develop the practical bandwidth selection rules for the two-step nonparametric estimators under the GPV specification, i.e., the specification proposed in Guerre et al. (2000); in section 4, we propose the conditional hazard rate based nonparametric estimator and develop its bandwidth selection rules; section 5 compares different nonparametric estimators by using a simulated procurement auction model and establish that the conditional hazard rate based nonparametric estimator with the proposed bandwidth selection rules is the best in finite sample estimation by the criterion of minimizing mean squared errors; in section 6, we apply the proposed conditional hazard rate based nonparametric estimator with its bandwidth selection rule to an empirical analysis of the Ohio highway procurement auction market; section 7 concludes.
2.2 The Procurement Auction Model with Heterogeneous Firms

A single and indivisible project is auctioned. All potential firms submit their bids simultaneously. The firm with the lowest bid wins the contract. It is assumed that there are two types of firms in the market: \( N_1 \geq 1 \) firms are of type 1 and \( N_2 \geq 1 \) firms are of type 2\(^7\). Each firm observes its private project cost to finish the project and does not know the project costs and bids of any other bidding firms. But it is common knowledge among all firms that the project cost of any type 1 firm is drawn independently from the distribution \( F_1(\cdot) \), which is absolutely continuous with the density function \( f_1(\cdot) \) on the support \([\underline{c}, \overline{c}]\) and that the project cost of any type 2 firm is drawn independently from the distribution \( F_2(\cdot) \), which is absolutely continuous with the density function \( f_2(\cdot) \) on the same support. The number of bidding firms and the type of each bidding firm are also common knowledge. The type-dependent symmetric BNE strategies of the game can be characterized by considering a representative firm for each type. As usual, we restrict the BNE strategies to be strictly increasing and differentiable. Let \( c_{1,i} \) and \( c_{2,i} \) denote the project costs for a type 1 firm \( i \) and a type 2 firm \( i \). Suppose the equilibrium strategies for type 1 and type 2 firms are \( s_1(\cdot) \) and \( s_2(\cdot) \). The type 1 firm \( i \) chooses its bid \( b_{1,i} \) to maximize its expected payoff, which can be expressed as:

\[
\max_{b_{1,i}} (b_{1,i} - c_{1,i}) \left[ 1 - F_1(s_1^{-1}(b_{1,i})) \right]^{N_1-1} \left[ 1 - F_2(s_2^{-1}(b_{1,i})) \right]^{N_2}
\]

\(^7\)When \( N_1 = 0, N_2 \geq 2 \) or \( N_1 \geq 2, N_2 = 0 \), the game is reduced to the symmetric game as in Guerre et al (2000). We choose two types of firms for the purpose of empirical application. The analysis can be extended to more than two types of firms.
The type 2 firm $i$ chooses its bid $b_{2,i}$ to maximize its expected payoff, which can be expressed as:

$$\max_{b_{2,i}} (b_{2,i} - c_{2,i}) * \left[ 1 - F_1(s_1^{-1}(b_{2,i})) \right]^{N_1} * \left[ 1 - F_2(s_2^{-1}(b_{2,i})) \right]^{N_2-1}$$

The first-order conditions of these expected payoff maximization problems form a system of two differential equations:

$$b_{1,i} = c_{1,i} + \frac{1}{(N_1 - 1) \frac{f_1(s_1^{-1}(b_{1,i}))}{1 - F_1(s_1^{-1}(b_{1,i}))} \frac{1}{s_1'(s_1^{-1}(b_{1,i}))} + \frac{2}{1 - F_2(s_2^{-1}(b_{1,i}))} \frac{1}{s_2'(s_2^{-1}(b_{1,i}))}}$$

$$b_{2,i} = c_{2,i} + \frac{1}{N_1 \frac{f_1(s_1^{-1}(b_{2,i}))}{1 - F_1(s_1^{-1}(b_{2,i}))} \frac{1}{s_1'(s_1^{-1}(b_{2,i}))} + \frac{2}{1 - F_2(s_2^{-1}(b_{2,i}))} \frac{1}{s_2'(s_2^{-1}(b_{2,i}))}}$$

for $i = 1, \cdots, N_1$ for the type 1 firm and $i = 1, \cdots, N_2$ for the type 2 firm.

Under the boundary conditions that $s_1(\bar{c}) = s_2(\bar{c}) = \bar{c}$, $s_1(\bar{c}) = s_2(\bar{c})$ and the relevant regularity conditions, the existence and uniqueness of the BNE strategies have been proved by Maskin and Riley (1996, 2000), LeBrun (1996, 1999) and Bajari (2001). The BNE bidding functions $s_1(\cdot)$ and $s_2(\cdot)$ in this asymmetric auction model generally do not have explicitly analytical expressions and are difficult to analyze directly. But following Guerre et al. (2000), we can use the distributions of bids of both types of bidders to obtain the inverse bidding functions. Let $G_1(b)$ and $G_2(b)$ denote the equilibrium distribution functions of the bids by type 1 firm and type 2 firm respectively, with the corresponding density functions $g_1(b)$ and $g_2(b)$. Using the facts that $b_1 = s_1(c_1)$ and $b_2 = s_2(c_2)$, we can obtain the following relationship:
\[
\frac{f_1(s_1^{-1}(b))}{1 - F_1(s_1^{-1}(b)) s_1'(s_1^{-1}(b))} = \frac{g_1(b)}{1 - G_1(b)} \quad (2.3)
\]

\[
\frac{f_2(s_2^{-1}(b))}{1 - F_2(s_2^{-1}(b)) s_2'(s_2^{-1}(b))} = \frac{g_2(b)}{1 - G_2(b)} \quad (2.4)
\]

By substituting equations (2.3) and (2.4) into equations (2.1) and (2.2), the system of two differential equations above can be transformed to

\[
c_{1,i} = b_{1,i} - \frac{1}{(N_1 - 1)(1 - G_1(b_{1,i})) + N_2 1 - G_2(b_{1,i})} g_1(b_{1,i}) \quad (2.5)
\]

\[
c_{2,i} = b_{2,i} - \frac{1}{N_1 (1 - G_1(b_{2,i})) + (N_2 - 1)(1 - G_2(b_{2,i}))} g_2(b_{2,i}) \quad (2.6)
\]

Equations (2.5) and (2.6) are the inverse bidding functions for type 1 and type 2 firms\(^8\).

Heterogeneity of project size usually exists in the field auction data and thus becomes a cost shifter. We assume that there exists an observed variable \(X\) that characterizes the project size. Then for the project \(l\) where \(X = x_l\), equations (2.5) and (2.6) will be written as

\[
c_{1,i,l} = b_{1,i,l} - \frac{1}{(N_1 - 1)(1 - G_1(b_{1,i,l}|x_l)) + N_2 1 - G_2(b_{1,i,l}|x_l)} g_1(b_{1,i,l}|x_l) \equiv \xi_1(b_{1,i,l}, x_l, N_1, N_2, \lambda_1, \lambda_2) \quad (2.7)
\]

\[
c_{2,i,l} = b_{2,i,l} - \frac{1}{N_1 (1 - G_1(b_{2,i,l}|x_l)) + (N_2 - 1)(1 - G_2(b_{2,i,l}|x_l))} g_2(b_{2,i,l}|x_l) \equiv \xi_2(b_{2,i,l}, x_l, N_1, N_2, \lambda_1, \lambda_2) \quad (2.8)
\]

\(^8\text{Bajari and Ye (2003) established the identification result for the general } N\text{-type asymmetric model.}\)
where

\[
\lambda_1(b|x) = \frac{g_1(b|x)}{1 - G_1(b|x)}
\]

\[
\lambda_2(b|x) = \frac{g_2(b|x)}{1 - G_2(b|x)}
\]

It is worth attention that the conditional hazard rate functions \( \lambda_1(b|x) \) and \( \lambda_2(b|x) \) depend on \( N_1 \) and \( N_2 \) because the bidding strategies are functions of \( N_1 \) and \( N_2 \) so that the distribution of equilibrium bids \( b_1 \) and \( b_2 \) depend on \( N_1 \) and \( N_2 \) as well. Guerre et al. (2000) proposed to include the number of bidders as another dimension of the conditioning variables\(^9\). However, this approach will place more strict requirements on the number of observations needed to obtain a good nonparametric estimation result. This data requirement is usually not satisfied in practice, especially when the asymmetric types of bidders are considered. We will estimate the conditional hazard rate function \( \lambda_1(b|x) \) and \( \lambda_2(b|x) \) for each given numbers of bidders \( N_1 \) and \( N_2 \) instead. Under the assumption that the underlying project cost distributions do not depend on the number of bidding firms, we can pool the project cost estimates for each pair of \( \{N_1, N_2\} \) in the second-step estimation. The estimation, a two-step procedure, follows easily from equations (2.7) and (2.8). At the first step, we use the data

\[
\{(b_{1,i,l}, x_l)_{i=1}^{N_1} \text{ and } (b_{2,i,l}, x_l)_{i=1}^{N_2} \text{ for } l = 1, \cdots, L_{N_1N_2}\}
\]

from \( L_{N_1N_2} \) auctions with \( N_1 \) type 1 firms and \( N_2 \) type 2 firms to estimate the pseudo samples of project costs:

\(^9\)Thus the conditioning variables become mixed, including both continuous and discrete variables.
At the second step we estimate the conditional density functions \( f_1(c|x) \) and \( f_2(c|x) \) by using the data \((\hat{c}_{1,i,l}, x_l)_{i=1}^{N_{1l}}\) and \((\hat{c}_{2,i,l}, x_l)_{i=1}^{N_{2l}}\) for \( l = 1, \ldots, L \).

2.3 Bandwidth Selection under GPV Nonparametric Specification

The two-step estimation method in the previous section originate from Guerre et al. (2000). The most important step is to estimate the inverse bidding function \( \xi_1(b_{1,i,l}, x_l, N_1, N_2, \hat{\lambda}_1, \hat{\lambda}_2) \) and \( \xi_2(b_{2,i,l}, x_l, N_1, N_2, \hat{\lambda}_1, \hat{\lambda}_2) \). In this section, we derive the practical bandwidth selection rules for the kernel nonparametric estimators under the specification in Guerre et al. (2000). For the type \( k \) firm \((k = 1\text{ and } 2)\), the conditional density function \( g_k(b|x) \) and conditional distribution function \( G_k(b|x) \) are estimated separately. This is the main characteristic of GPV specification, as different from the conditional hazard rate based specification proposed in the next section. The notations below are used to be consistent with the auction model. Assume that we have \( L \) independent observations of \( x \) playing the role of cost shifter, denoted by \( \{x_l\}_{l=1}^L \). Given any \( x_l \), we have \( N_k \) observations of \( b \) resulting from type \( k \) firms independent equilibrium bids, denoted by \( \{b_{k,i,l}\}_{i=1}^{N_{kl}} \). The estimation strategy under GPV specification is as follows. In the first-step, the conditional density and distribution functions \( g_k(b|x) \) and \( G_k(b|x) \) are estimated separately by:
\[
\hat{g}_k(b|x) = \frac{\hat{g}_k(b, x)}{\hat{f}(x)} \quad \text{and} \quad \hat{G}_k(b|x) = \frac{\hat{G}_k(b, x)}{\hat{f}(x)}
\]

where

\[
\hat{g}_k(b, x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{N_k h_g} \sum_{i=1}^{N_k} k\left( \frac{b - b_{k,i,l}}{h_g} \right) k\left( \frac{x - x_l}{h_g} \right) \tag{2.9}
\]

\[
\hat{G}_k(b, x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{N_k h_G} \sum_{i=1}^{N_k} 1(b_{il} \leq b) k\left( \frac{x - x_l}{h_G} \right) \tag{2.10}
\]

and

\[
\hat{f}(x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{h_x} k\left( \frac{x - x_l}{h_x} \right)
\]

In the second step, the conditional density function of project cost for type \( k \) firm \( f_k(c|x) \) is estimated by

\[
\hat{f}_k(c|x) = \frac{\hat{f}_k(c, x)}{\hat{f}(x)}
\]

where

\[
\hat{f}_k(c, x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{N_k h_f} \sum_{i=1}^{N_k} k\left( \frac{c - \hat{c}_{k,i,l}}{h_f} \right) k\left( \frac{x - x_l}{h_f} \right)
\]

\[
\hat{c}_{1,i,l} = b_{1,i,l} - \frac{1}{(N_1 - 1) \hat{g}_1(b_{1,i,l}|x_l) + N_2 \hat{g}_2(b_{1,i,l}|x_l) + 1 - G_1(\hat{c}_{k,i,l}|x_l)} \tag{2.11}
\]

and

\[
\hat{c}_{2,i,l} = b_{2,i,l} - \frac{1}{N_1 \hat{g}_1(b_{2,i,l}|x_l) + (N_2 - 1) \hat{g}_2(b_{2,i,l}|x_l) + 1 - G_2(b_{2,i,l}|x_l)} \tag{2.12}
\]
As shown by Guerre et al. (2000), given the theoretical bandwidth choices:

\[
\begin{align*}
    h_g &= \lambda_g (\log L/L)^{1/(2R+3)}, \quad h_G = \lambda_G (\log L/L)^{1/(2R+4)} \\
    h_f &= \lambda_f (\log L/L)^{1/(2R+4)}, \quad h_x = \lambda_x (\log L/L)^{1/(2R+3)}
\end{align*}
\]

\(\hat{f}(\cdot|\cdot)\) will uniformly converge to \(f(\cdot|\cdot)\) at the optimal rate \((L/\log L)^{R/(R+4)}\). However, the theoretical bandwidths above fall short of providing a practical guidance for the empirical application since they include the unknown constants \(\lambda_g\), \(\lambda_G\), \(\lambda_f\) and \(\lambda_x\). In applying the nonparametric method in the empirical analysis, the literature has adopted the following rule-of-thumb as the bandwidth choice rules under the assumption \(R = 1\):

\[
\begin{align*}
    h_g &= c_g (NL)^{-1/6}, \quad h_f = c_f (NL)^{-1/6} \\
    h_G &= c_G (NL)^{-1/5}, \quad h_x = c_x L^{-1/5}
\end{align*}
\]

where \(c_g = 2.978 \times 1.06 \times \hat{\sigma}_b\), \(c_G = 2.978 \times 1.06 \times \hat{\sigma}_b\), \(c_f = 2.978 \times 1.06 \times \hat{\sigma}_c\), \(c_x = 2.978 \times 1.06 \times \hat{\sigma}_x\) and \(\hat{\sigma}_b\), \(\hat{\sigma}_x\) and \(\hat{\sigma}_c\) are the standard deviations of bids \(\{b_{k,i,l}\}_{i=1}^{N_k} \in L\) project sizes \(\{x_l\}_{l=1}^L\) and pseudo sample costs \(\{{\hat{c}}_{k,i,l}\}_{i=1}^{N_k} \in L\). Using this rule-of-thumb, Li and Perrigne (2003) estimated the timber sale auction where the appraisal value of the timer for sale characterized the heterogeneity in auction size. Flambard and Perrigne (2004) estimated asymmetric snow removal procurement auctions where a constructed variable characterized the heterogeneity in auction size. However, several practical problems exist for these commonly used bandwidth rules.

\(^{10}\) \(F(\cdot|\cdot)\) and \(f(\cdot|\cdot)\) are assumed to admit up to \(R + 1\) continuous bounded partial derivatives.

\(^{11}\) 2.978 follows from the regular use of triweight kernel instead of a Gaussian kernel.
First, the constants are calculated using the total variation of the corresponding variables in the data such as all bids, which will result in oversmoothing. In estimating a density or distribution function conditional on $X = x$, only local observations around the conditioning variable $x$ can provide useful information for estimation. Second, the estimated conditional bid density function $\hat{g}(b|x)$ tends to be multi-modal. Third, the estimated conditional bid distribution function $\hat{G}(b|x)$ is not restricted to the interval $[0, 1]$, and thus the estimated project cost can be higher than the observed bid. The reason for the latter two problems is that the denominator and numerator in each estimators involve different bandwidth choices. To overcome these practical problems, we will base on the joint density function $g(b, x)$ to derive practical rule $h_g$ and base on the conditional distribution function $G(b|x)$ to derive practical rule $h_G$ with the restriction that $h_x = h_G$, which guarantees that the conditional distribution function $\hat{G}(b|x)$ is restricted to the interval $[0, 1]$ without losing its asymptotic optimality. The proposition below provides guidance for choosing bandwidths for estimating $\hat{g}_k(b|x)$ and $\hat{G}_k(b|x)$.

**Proposition 2** Under assumptions (1)-(4) in the appendix, define the kernel estimators of the joint density function $f(x, y)$ and conditional distribution function $F(y|x)$ as follows:

\[
\hat{f}(y, x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_f^2} \sum_{i=1}^{N} k\left(\frac{y - y_{il}}{h_f}\right) k\left(\frac{x - x_l}{h_f}\right)
\]

\[
\hat{F}(y|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_F} \sum_{i=1}^{N} 1(y_{il} \leq y) k\left(\frac{x - x_l}{h_F}\right)
\]

\[
\frac{1}{L} \sum_{l=1}^{L} \frac{1}{h_F} k\left(\frac{x - x_l}{h_F}\right)
\]
Then weighted IMSE for \( \hat{f}(y, x) \) and \( \hat{F}(y|x) \) can be expressed as follows:

\[
IMSE(\hat{f}(y, x)) = \frac{c_{11}}{Lh_f^2} + c_{12}h_f^4, \quad IMSE(\hat{F}(y|x)) = \frac{c_{21}}{Lh_F} + c_{22}h_F^4,
\]

where \( c_{11}, c_{12}, c_{21} \) and \( c_{22} \) are constants\(^{12}\). The optimal bandwidths minimizing IMSE are

\[
h_f^* = \left( \frac{2c_{12}}{c_{11}} \right)^{-1/6} L^{-1/6} \quad \text{and} \quad h_F^* = \left( \frac{4c_{22}}{c_{21}} \right)^{-1/5} L^{-1/5}.
\]

The proposition above describes the optimal bandwidth selection rules for estimating the joint density function \( g(b, x) \) and the conditional distribution function \( G(b|x) \) in the auction setting. However they are still not practical because the constants \( c_{11}, c_{12}, c_{21} \) and \( c_{22} \) depend on the unknown true conditional density function, the unknown true conditional distribution function and the unknown true marginal density function. Like the Silverman’s rule of thumb, we can assume the form of the underlying distributions and the way in which \( y \) depends on \( x \) so as to obtain plug-in bandwidth selection rules. In practice we use the normal reference distribution and the linear functional form, i.e., \( y = \beta_0 + \beta_1 x + e \) where \( e \) follows a normal distribution. Then we can calculate the constants \( c_{11}, c_{12}, c_{21} \) and \( c_{22} \), from which the optimal bandwidths follow\(^{13}\). We can expect this method to work well even if the underlying distribution is not normal. Numerical integration method is used in calculating \( c_{11}, c_{12}, c_{21} \) and \( c_{22} \).

The proposition above gives the practical bandwidth selection rules for choosing \( h_g, h_G, h_f \) and \( h_x \) in the GPV specification. Asymptotically, the proposed bandwidths

\(^{12}\)The expressions for \( c_{11}, c_{12}, c_{21} \) and \( c_{22} \) are given in the appendix.

\(^{13}\)\( \beta_0, \beta_1 \) and the parameters in the normal reference distribution can be estimated by using OLS or MLE on the observed data.
are equivalent to the commonly adopted bandwidths. But they could be quite different in finite sample estimation. For example, unlike the usual choice \( h_x \) that involves the total variation of \( \{x_i\}_{i=1}^L \), our choice of \( h_x \) only involves the local variation around the conditioning variable \( x \)\(^{14}\).

### 2.4 Bandwidth Selection Under Conditional Hazard Rate Based Nonparametric Specification

Under the GPV specification in the previous section, the bandwidths in the estimators of the joint density and conditional distribution function are derived separately. However, our purpose is to estimate the inverse bidding function rather than estimate the joint density or conditional distribution functions, we can expect to obtain better smoothing results by jointly considering the bandwidth selection. Note that the inverse bidding function can be written as follows:

\[
\begin{align*}
\xi_1(b, x, N_1, N_2, \lambda_1, \lambda_2) &= b - \frac{1}{(N_1 - 1)\lambda_1(b|x) + N_2\lambda_2(b|x)} \\
\xi_2(b, x, N_1, N_2, \lambda_1, \lambda_2) &= b - \frac{1}{N_1\lambda_1(b|x) + (N_2 - 1)\lambda_2(b|x)}
\end{align*}
\]

where

\[
\begin{align*}
\lambda_1(b|x) &= \frac{g_1(b|x)}{1 - G_1(b|x)} \\
\lambda_2(b|x) &= \frac{g_2(b|x)}{1 - G_2(b|x)}
\end{align*}
\]

\(^{14}\)In principle, we can derive a different bandwidth rule \( h_x \) in estimating the conditional density functions \( g(b|x) \) and \( f(c|x) \) rather than use the same one as used in estimating the conditional dis tribution function \( G(b|x) \), which will result in smaller MSE. We choose to use the same \( h_x \) in order to be consistent with the bandwidth that is proposed by GPV and commonly used in empirical literature.
which means that the inverse bidding functions depend on the conditional hazard rate functions \( \lambda_1(b|x) \) and \( \lambda_2(b|x) \). We propose to estimate \( \lambda_k(b|x) \) directly rather than estimate \( G_k(b|x) \) and \( g_k(b|x) \) separately for \( k = 1 \) and \( 2 \).

The conditional density function \( g_k(b|x) \) can be estimated by

\[
\hat{g}_k(b|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_x} \sum_{i=1}^{N_k} k\left(\frac{b-b_{k,i,l}}{h_b}\right)k\left(\frac{x-x_i}{h_x}\right)
\]

(2.13)

This is similar to the kernel conditional density estimator in Hyndman (1996). Then a natural estimate for \( G(b|x) \) is given by

\[
\hat{G}_k(b|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_x} \sum_{i=1}^{N_k} K\left(\frac{b-b_{k,i,l}}{h_b}\right)k\left(\frac{x-x_i}{h_x}\right)
\]

(2.14)

where the properties of the functions \( k(\cdot) \) and \( K(\cdot) \) are given in the appendix. The specification for \( \hat{G}_k(b|x) \) above guarantees that it is restricted to \([0, 1]\). From equations (2.13) and (2.14) we can estimate the conditional hazard rate function \( \lambda_k(b|x) \) by

\[
\hat{\lambda}_k(b|x) \equiv \frac{\hat{g}_k(b|x)}{1 - \hat{G}_k(b|x)} = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_x} \sum_{i=1}^{N_k} k\left(\frac{b-b_{k,i,l}}{h_b}\right)k\left(\frac{x-x_i}{h_x}\right)
\]

and then estimate the unobserved project cost by

\[
\hat{c}_{k,i,l} = \xi_k(b_{k,i,l}, x_l, N_1, N_2, \hat{\lambda}_1, \hat{\lambda}_2)
\]

(2.15)

The following proposition shows the asymptotic bias and variance of the conditional hazard rate estimator \( \hat{\lambda}_k(b|x) \).
Proposition 3  Given the assumptions (1)-(4) in the appendix and
\( Lh_y \rightarrow \infty, h_y \rightarrow 0, h_x \rightarrow 0 \) as \( L \rightarrow \infty \), the kernel estimator of the conditional hazard rate
function defined by

\[
\hat{\lambda}(y|x) = \frac{\hat{f}(y|x)}{1 - F(y|x)} = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_y h_x} \sum_{i=1}^{N} k\left(\frac{y_i - y}{h_y}\right) k\left(\frac{x - x_i}{h_x}\right)
\]

satisfies

\[
\sqrt{Lh_y h_x} \left[ \hat{\lambda}(y|x) - \lambda(y|x) - \text{AsyBias}[\hat{\lambda}(y|x)] \right] \overset{d}{\rightarrow} N(0, \frac{R^2_0(k^2)g(y|x)}{Nh(x)S^2(y|x)})
\]

where

\[
\text{AsyBias}[\hat{\lambda}(y|x)] = \frac{\sigma^2_k}{2} \left[ \frac{\partial^2}{\partial y^2} \left[ f(y|x)h(x) \right] - \frac{\partial^2}{\partial y^2} \left[ S(y|x)h(x) \right]f(y|x) \right] h^2_x
\]

\[
+ \frac{\sigma^2_k}{2} \left[ \frac{f''(y|x)}{S(y|x)} + \frac{f'(y|x)f(y|x)}{S^2(y|x)} \right] h^2_y + O(h^4_x) + O(h^2_x h^2_y) + O(h^4_y)
\]

where \( S(y|x) = 1 - F(y|x) \).

It follows that the asymptotic mean square error of \( \hat{\lambda}(b|x) \) at \( (x, b) \) is

\[
AMSE(x, b) = \left[ E\left[ \hat{\lambda}(b|x) \right] - \lambda(b|x) \right]^2 + Var[\hat{\lambda}(b|x)]
\]

\[
= C_1(x, b) \frac{1}{Lh_x h_b} + C_2(x, b)h^4_x + C_3(x, b)h^4_b + C_4(x, b)h^2_x h^2_b
\]

where \( C_1(x, b), C_2(x, b), C_3(x, b) \) and \( C_4(x, b) \)\(^\text{15}\) are constants depending on the kernel
function \( k(\cdot) \), the conditional density function \( g(y|x) \), the conditional distribution
function \( G(y|x) \) and the marginal density \( h(x) \). In order to obtain a bandwidth
choice rule which depends on the conditioning variable \( x \) rather than a uniform rule,

\(^{15}\)Their expressions are in the appendix.
we choose to minimize the integrated asymptotic mean square error conditional on \( x \) as the criterion to decide the bandwidth \( h_b(x) \) and \( h_x(x) \). This approach will result in smaller bandwidth in regions with more data and larger bandwidth in regions with less data and thus is more practical. We integrate \( AMSE(x, b) \) over the domain of \( b \) to obtain the \( IMSE(x) \):

\[
IMSE(x) = \int AMSE(x, b) db
\]

\[
= \frac{C_1(x)}{L} + C_2(x)h_x^4 + C_3(x)h_b^4 + C_4(x)h_x^2h_b^2
\]

where

\[
C_1(x) = \int C_1(x, b) db, \quad C_2(x) = \int C_2(x, b) db
\]

\[
C_3(x) = \int C_3(x, b) db, \quad C_4(x) = \int C_4(x, b) db
\]

Following Hyndman et al. (1996), the optimal bandwidth \( h_x^*(x) \) and \( h_b^*(x) \) are as follows:

\[
h_x^*(x) = C_1^{1/6}(x) \left[ 4 \left( \frac{C_2(x)}{C_3(x)} \right)^{1/4} + 2C_4(x) \left( \frac{C_2(x)}{C_3(x)} \right)^{3/4} \right]^{-1/6} L^{-1/6} \tag{2.16}
\]

\[
h_b^*(x) = C_1^{1/6}(x) \left[ 4 \left( \frac{C_2(x)}{C_3(x)} \right)^{1/4} + 2C_4(x) \left( \frac{C_2(x)}{C_3(x)} \right)^{3/4} \right]^{-1/6} L^{-1/6} \tag{2.17}
\]

The constants \( C_1(x), C_2(x), C_3(x) \) and \( C_4(x) \) can be explicitly calculated so as to obtain the plug-in bandwidth selection rules by assuming the form in which \( y \) depends on \( x \) and the form of the underlying distributions.

In the second step, we estimate the conditional density function \( f(c|x) \) as usual by
The bandwidth choice rules can be obtained using the method introduced in the previous section\textsuperscript{16}.

2.5 Comparison between Different Methods and Monte Carlo Experiment

We have presented three nonparametric estimation methods in terms of different kernel-based specifications or bandwidth selection rules. Suppose method 0 uses the commonly adopted bandwidth rules under the GPV specification, method 1 uses the proposed bandwidth rules in section 3 under the GPV specification and method 2 uses the proposed bandwidth rules in section 4 under the specification based on conditional hazard rate. We will compare the inverse bidding function estimators under these three different methods. Since the target density function $f(\cdot | \cdot)$ is estimated using the pseudo sample costs obtained from the estimated inverse bidding function, a good estimate of the inverse bidding function is critical to obtain a good estimate of the target density function. Let $\xi(b|x)$ and $\hat{\xi}(b|x)$ denote the true and estimated inverse bidding functions conditional on the project size $x$. First we consider the conditional integrated mean square error of $\hat{\xi}(\cdot|x)$ defined by

$$CIMSE(\hat{\xi}, \xi|x) = \int E[\hat{\xi}(b|x) - \xi(b|x)]^2 db$$

\textsuperscript{16}We can also estimate the conditional density function as that in Hyndman et al. (1996), i.e.,

$$\hat{f}_k(c|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_f} \sum_{i=1}^{N} k(\frac{c-\hat{\xi}_{k;i;l}}{h_f})k(\frac{x-x_i}{h_x})$$

$$\hat{f}_k(c|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_x} k(\frac{x-x_i}{h_x})$$

$$\hat{f}_k(c|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_x} k(\frac{x-x_i}{h_x})$$
which can be used to evaluate the inverse bidding function estimator \( \hat{\xi}(\cdot|x) = \hat{\xi}(\cdot, data|x) \).

Second we consider the weighted integrated mean square error of \( \hat{\xi}(\cdot) \) defined by

\[
IMSE(\hat{\xi}, \xi) = \int \int [\hat{\xi}(b|x) - \xi(b|x)]^2 h(x) db dx
\]

which can be used to evaluate the overall performance of inverse bidding function estimator \( \hat{\xi}(\cdot) = \hat{\xi}(\cdot, data|\cdot) \) over the entire domain of \( x \), with more emphasis on the region with more data. For the numerical example below, we will estimate \( CIMSE(\hat{\xi}, \xi|x) \) by

\[
CIMSE(\hat{\xi}, \xi|x) = \sum_{s=1}^{S} \sum_{j=1}^{M1} [\hat{\xi}(b^x_j, data(s)|x) - \xi(b^x_j|x)]^2
\]

and estimate \( IMSE(\hat{\xi}, \xi) \) by

\[
IMSE(\hat{\xi}, \xi) = \frac{\Delta x}{SL} \sum_{s=1}^{S} \sum_{j=1}^{M2} \sum_{l=1}^{L} [\hat{\xi}(b^x_j, data(s)|x_l) - \xi(b^x_j|x_l)]^2
\]

where \( data(s) = \{x_l, b_{i,l}|l = 1, \ldots, L; i = 1, \ldots, N\} \) is a random sample from \( L \) independent heterogeneous auctions with \( N \) bidders, \( b^x = \{b^x_1, \ldots, b^x_{M1}\} \) is a vector of \( M1 \) equally spaced values over the domain of \( b \) conditional on \( x \) with \( b^x_{j+1} - b^x_{j} = \Delta x \), \( b = \{b_1, \ldots, b_{M2}\} \) is a vector of \( M2 \) equally spaced values over domain of \( b \) with \( b_{j+1} - b_j = \Delta \), and \( \hat{\xi}(b^x_j, data(s)) \) is calculated from \( data(s) \) using (2.15). By simulating \( CIMSE \) and \( IMSE \) under each method, we can compare the performance of different inverse bidding function estimators.

We use the three methods discussed above to estimate a simulated procurement auction model. The setting is a symmetric low-price sealed-bid private cost procurement auction. The Monte Carlo experiments consist of 50 replications. In each
replication, assume there are 200 auctions in the data. For each auction, 5 firms compete for an indivisible project. The project size $x$ is drawn from a normal distribution with mean 10 and standard deviation 3, truncated on $[4, 16]$. For a given project size $x$, the costs for all 5 firms are drawn identically and independently from a normal distribution with mean $0.2 + 0.8x$ and standard deviation 0.6, truncated on $[-1.6 + 0.8x, 2 + 0.8x]$. Given the assumptions above, the bidding function $s(c|x, F)$ is given by

$$b = s(c|x, F) = c + \int_{c}^{c+0.8x} \frac{[1 - F(u|x)]^{N-1} du}{[1 - F(c|x)]^{N-1}}$$

(2.20)

where $N = 5$ and $F(\cdot | \cdot)$ is the relevant truncated conditional normal distribution. The data generation process goes as follows: first, draw 200 i.i.d. project sizes; second, draw 5 i.i.d. project costs for each of the 200 project sizes; third, calculate 5 bids in each project by using bidding function (2.20). In the third step the Quasi-Monte Carlo method is used to calculate the integration in the bidding function numerically.

Then we estimate the conditional inverse bidding function and the conditional project cost density function by different methods. For all three methods, two practical issues are worthy of attention. First, it is well-known that the nonparametric kernel method will have biased estimates near the boundary. This means the estimated pseudo-costs corresponding to the bids near the boundary are biased. Following Guerre et al. (2000), we use the following simple trimming rules: for each $x$, define a half-closed and half-open interval centered around $x$ with width $d$ as $I(x) \equiv [x - d, x + d)$\textsuperscript{17} and

\textsuperscript{17}If $x + d$ is larger than or equal to the maximum of the observed $x$, $x_{\text{max}}$, the interval would be $[x - d, x_{\text{max}}]$.
\[ b_{\min}(x) = \min \{ b_{i,l}, i = 1, \cdots, N \} \text{ for } \forall l \text{ s.t. } x_l \in I(x) \]

\[ b_{\max}(x) = \max \{ b_{i,l}, i = 1, \cdots, N \} \text{ for } \forall l \text{ s.t. } x_l \in I(x) \]

\[ x_{\max} = \max \{ x_l, l = 1, \cdots, L \}, \quad x_{\min} = \min \{ x_l, l = 1, \cdots, L \} \]

The pseudo-sample cost would then be trimmed by the following rule:

\[ \hat{c}_{il} = \left\{ \begin{array}{ll}
\hat{\xi}(b_{il}, x_l, n, \lambda) & \text{if } b_{\min}(x_l) + h_g \leq b_{il} \leq b_{\max}(x_l) - h_g \\
+\infty & \text{and } x_{\min} + h_x \leq x_l \leq x_{\max} - h_x
\end{array} \right. \]

for \( i = 1, \cdots, 5 \) and \( l = 1, \cdots, 200 \). Positive infinity is assigned to the value of costs corresponding to the bids near the boundary, which has no effect in estimating the conditional density function of project cost. In our simulation study \( d = \max(h_g, h_x) \).

Second, if we assume the density function of project cost is symmetric, the distribution of the bids will be skewed to the right, which results from the fact that the bidding function is above and approaching the 45-degree line as the cost increases. The skewness of the bids is usually observed in the field data. If there exists high skewness in the observed bid, we need to transform the bid to decrease the effect of skewness on the nonparametric estimates. Define \( d_{i,l} = \log(1 + b_{i,l}) \) and \( z_l = \log(1 + x_l) \). Then the inverse bidding function can be written as

\[ c_{i,l} = \exp(d_{i,l}) - 1 - \frac{\exp(d_{i,l})}{(N-1)\lambda(d_{i,l}|z_l)} \]

All the procedures in the first step to recover pseudo sample cost are the same as before after we transform the data, except that we are now estimating the hazard rate function of \( d \) conditional on \( z \).

We plot the estimates of the inverse bidding functions corresponding to \( x = 8, 10 \) and 12 in Figure 1 in the appendix. The first, second and third rows are obtained
by the methods 0, 1 and 2 respectively. The first, second and third columns correspond to $x = 8, 10$ and 12 respectively. The inverse bidding functions are evaluated at 30 equally spaced points on their relevant supports for each project size. At each point, the mean, 5th percentile and 95th percentile of the 50 estimates are displayed. The solid line is the real inverse bidding function. The dotted lines are 5th percentile, mean and 95th percentile of the estimated inverse bidding functions. It is easy to see that under the method 0, the real inverse bidding function falls outside of the 90% confidence interval on a large part of bid support. In fact the commonly adopted bandwidth will result in underestimated cost on average. In contrast, the methods 1 and 2 produce very good estimates of the inverse bidding functions. The real inverse bidding function falls in the 90% confidence interval and almost coincides with the mean of the estimated inverse bidding functions. Table 1 in the appendix displays the bandwidths used in simulating the inverse bidding function under these three methods for $x = 10$. Table 2 in the appendix reports the estimated $CIMSE$ and $IMSE$ defined by (2.18) and (2.19) corresponding to $x = 8, 10$ and 12. The methods 1 and 2 result in much lower $IMSE$ or $CIMSE$ than the method 0, with method 2 more preferred. Due to the boundary effect, the estimated inverse bidding functions by the methods 1 and 2 have relatively large variance near the lower bound. This is similar to the situation in Guerre et al. (2000) where a simulated auction model without heterogeneity is estimated. Next we plot the estimates of the density function of project cost conditional on $x = 10$ in Figure 2 in the appendix. The first and second columns are obtained by the methods 0 and 2 respectively. On average the proposed conditional hazard rate based estimator with its corresponding bandwidths produces good estimates of the conditional density functions and the real conditional
density functions fall in the 90% confidence interval, but the conditional density function estimated by commonly used bandwidths is excessively spread. Note that it is not unexpected the proposed method gives a relatively wider confidence interval in estimating the conditional density function because the criterion of minimizing $IMSE$ is based on a trade-off between the bias and the variance.

From Table 1 we can see that the most apparent difference among these bandwidth choices is that the commonly used bandwidth is much larger than the proposed bandwidth. The former uses the total variation of the relevant variables in the data. When the observed data is characterized with heterogeneity where the conditional variance is usually much smaller than the total variance, the bandwidth will be overestimated by the commonly used method. The overestimated bandwidth in finite sample estimation will result in biased project cost estimates and more widely spread density function estimates. The overestimated bandwidth can also give a biased estimate for the mean of conditional project cost distribution. For small project sizes, we would expect the estimated conditional density function will be skewed to the right; for large project sizes, we would expect the estimated conditional density function will be skewed to the left.

It is worth mentioning that the commonly used bandwidth rules are applicable if no heterogeneity exists across different auctions. This can be seen from the simulation study in Guerre et al. (2000). In Campo, S. Perrigne, I. and Vuong, Q. (2003), they made the strong assumption that the heterogeneity across different projects was incorporated in the number of bidders, so they did not consider heterogeneity explicitly. In Li and Perrigne (2003), they estimated the conditional density functions of private
values in the first price sealed-bid auction. Their estimated conditional density function had a support not sensitive to the appraisal value, the variable characterizing the heterogeneity across the timber sale auctions. Their results are acceptable only if the support and the variance of the true conditional density function remain unchanged across different auctions. But in the low-price seal-bid auction such as the highway procurement auction, this is usually an unsatisfactory assumption. It is unrealistic to assume that a firm would form its own bid thinking that other rival firms could accomplish the project for half or an even smaller amount of its cost with a high probability when the project is relative large; it is also unrealistic to assume that a potential firm would form its own bid thinking that other rival firms would accomplish the project for double or an even larger amount of its cost with a high probability when the project is relatively small. In other words, the supports of the project costs and bids, and the distributions of project costs and bids depend on the project size. We need to take this dependence into account in choosing the bandwidths.

2.6 Application to the Ohio Highway Procurement Market

In this section we will study the Ohio highway procurement market using the kernel estimators under the conditional hazard rate based specification. More specifically, we will identify and estimate the potential asymmetry among the bidding firms in a structural way. The public procurement auction plays an important role in the U.S. economy. The state and local governments across the U.S. use procurement to administer public affairs such as highway construction and maintenance. In Ohio, the total value of the highway construction contracts awarded by the Ohio Department
of Transportation (ODOT) in 2002 was $1203 million, which accounted for 23% of Ohio’s total procurement expenditures in that year.

2.6.1 Description of the Ohio Highway Procurement Market

The Ohio Department of Transportation (ODOT) is the official government institution for constructing and maintaining the highways in Ohio. It usually invites tenders for highway-related projects twice a month and awards each project to one of the potential firms through a low-price sealed-bid auction. The project types include new construction, lane resurfacing, new bridge construction, bridge replacement, bridge repair, relocation, widening, painting, marking, signing, miscellaneous and etc. The awarding process is as follows. First, the ODOT advertises each identified project about 5 to 6 weeks before the letting date. Second, all potential firms pre-qualified by the ODOT purchase plans of the projects that they are interested in from ODOT, which provide firms with the documents containing information about the project such as the location and description of the project, an estimate of quantities and a description of the item-by-item work, the time to complete the work, the amount of the proposal guaranty, the department’s deadline for receiving a completed bid, and etc. Third, any bidding firm submits its sealed-bid to the ODOT before the deadline. Finally, all submitted bids are opened in public on the letting date and the project is awarded to the firm with the lowest bid.

No existence of binding reservation prices is usually justifiably assumed in the highway construction procurement auction and this assumption holds in the Ohio market. For each project, ODOT forms a department’s estimated cost, also called engineer’s estimate, which is based on engineers or experts’ estimation and previous
homogeneous projects. According to relevant regulations, if the department’s estimate for the cost is not confidential, the ODOT will not award a contract for an amount greater than 105% of the estimate. If the department’s estimate is confidential and the successful bid exceeds the estimate, the director needs to determine that the successful bid is fair and reasonable, and that the winner’s bid does not exceed the maximum permitted by the federal regulations, which requires that the winning bid should be lower than 110% of the department’s estimate and a higher successful bid than this threshold needs to be justified in writing. In the data about 28% of all bids are 5% higher than the engineer’s estimate, about 16% of all bids and 2% of all winning bids are 10% higher than the engineer’s estimates. These facts suggest that the bidders usually do not take the state and federal restrictions into account when determining their bids. So we will assume there is no effective reservation price in this market. We will also assume that the number of actual bidders is known to the participants in the auction, a typical assumption in the existing literature on highway construction procurement auctions.

2.6.2 The Data and Some Reduced-Form Analysis

The data we have for the highway procurement contracts awarded by ODOT are from January 1998 to December 2004. Among all available projects, we have focused on the two-lane and four-lane resurfacing projects since they are homogeneous, constitute the largest portion of all projects in terms of the quantity awarded, have the relatively simple work type such that firms can predict their cost very well and have relatively little uncertainty in the future construction. As indicated by Hong and Shum (2001), compared to other work types which may have common value feature
or both private and common value features, these projects are predominantly private value auctions. The total value of the 1259 contracts awarded in the data is $1.94 billion, which accounts for approximately 27% of the total value of all projects awarded by ODOT during the sample period. The geographical distribution of all projects and firms can be found in Figure 3 in the appendix. Next we analyze the observed heterogeneity of the highway construction projects and bidding firms.

**Heterogeneity across Projects** We observe heterogeneity across projects mainly in two dimensions in the data: the project size and the number of firms participating in the auction. The engineer's estimate could be the best single variable to characterize the project size because of its relatively consistent calculation method. All observed bids are distributed around the engineer’s estimates. As seen from Figure 4, the normalized bid (the bid divided by the engineer’s estimate) fluctuates with mean 0.9851 and standard deviation 0.1332. The normalized winning bid fluctuates with mean 0.9279 and standard deviation 0.1072. This can be further justified by the regression analysis shown in Table 7: the engineer’s estimate is extraordinarily significant in explaining the variation in the bids. We see from Table 4 that the engineer’s estimate ranges from $16,000 to $73.57 million with a mean $1.64 million and a standard deviation $3.658 million. In the structural estimation, we will use the engineer’s estimate to characterize the heterogeneity across projects. From Figure 5, we can see that the range of bids in an auction depends on the engineer’s estimate, and the variation of bids conditional on a given engineer’s estimate is much smaller than the total variation of bids and the total variation of engineer estimate. Thus the engineer’s estimate does play the role of cost shifter. If we assume that the support of the bids does not depend on the engineer’s estimate (i.e. we assume a much wider
support for the bids), then competition among firms would be strengthened when the project size is relatively large, and the competition among firms would be weakened when the project size is relatively small. The number of bidding firms in the data changes from project to project with mean 3 and standard deviation 1.56.

We assume that there is no unobserved heterogeneity across projects. In Krasnokutskay (2004), unobserved heterogeneity is identified in Michigan highway procurement auctions by observing that the regression analysis of the observed bids will improve a lot by including the project-specific dummies. Our regression result shows that the unobserved heterogeneity is not significant in the Ohio market. 98% of the total variations in the bids can be accounted for by the engineer’s estimate.

**Heterogeneity across Firms** As usual we assume that the only unobserved heterogeneity across firms is characterized by the unobserved project cost of each firm to accomplish a project. This is partly explained by the different bids submitted by the bidding firms. If we assume no observable heterogeneity, a symmetric independent private value (IPV) model then follows. However, in reality the bidding firms differ in terms of their size, location, specialization and frequency of bidding for projects of a specific type. These aspects imply the possible observable heterogeneity or asymmetry across firms. If the bidding firms take the asymmetry into account in deciding their bids, the symmetric IPV model would be inappropriate. In order to get more robust estimation result and policy guidance, we need to identify the possible ex ante asymmetries among firms that they take into account in submitting their bids and estimate their respective bidding strategies and project cost distributions.

The most striking asymmetry among firms in the data is their different frequencies of bidding. In the data a total of 157 firms participated in a total of 1259 projects.
Table 5 shows that most of the firms participated in less than 1% of all auctions and only less than 10 firms frequently bid for projects in this market. On the one hand, as indicated in Kranokutskaya (2004), firms that infrequently participate in the highway procurement market may have higher costs due to their limited experience with ODOT. On the other hand, firms that infrequently participation in the market for resurfacing projects may have some inherent cost or opportunity cost disadvantage in this type of work. This assumption is supported by the data. If firms are symmetric, we should expect that the frequency of bidding and the frequency of winning are comparable. From Table 5 we find that the firms that bid more frequently have a relatively higher frequency of winning. But this comparison may be inappropriate because some projects only had infrequently bidding firms or only had frequently bidding firms. We will give a more in-depth comparison below after we define two types of firms.

By the criterion of frequency of bidding we divide all firms into two types: regular firms and fringe firms. The regular firms are those that frequently participated in this market. Although such a division among firms may be parsimonious and more precise delineation can be given, this classification can characterize the main existing asymmetry among firms. Now we restrict the comparison to the samples with at least one regular firm and at least one fringe firm. There are 566 projects and 2138 bids in this subgroup. We further divide this subgroup into three categories by the engineer’s estimates of the projects as shown in Table 6. We calculate the mean of the probability of winning for fringe firms given the assumption that both regular firms and fringe firms are symmetric and the frequency of winning for fringe firms. Table 6 shows that it is more possible for a fringe firm to win when the project size
is relatively small and more possible for a regular firm to win when the project size is relatively large\textsuperscript{18}. Another observed asymmetry among firms is the distance between firms and the county where the project is located. Generally speaking, the fringe firms often participate in the local projects with an average working distance of 34 miles. The regular firms usually participate in the projects on a larger scope with an average working distance of 62 miles. Since firms need to move the required equipment to the project site and workers may need to commute between work site and home, that distance may affect both firms’ decisions to bid and the amount to bid. We can use OLS and Probit models to analyze the magnitude of the effect that the distance may have. The estimation results are shown in Table 7, and variables used in estimation are explained in Table 3. To explain the variation in the observed participation and bids, it is necessary to identify the project-specific, firm-specific and rival-specific characteristics that could explain those variations. For each project awarded, we have the following observations: the engineer’s estimate, the estimated working days, the bid by each participating firm, the distance of each firm to the project\textsuperscript{19} and the number of participating firms. The minimum distance to the project of the other participating rival firms is used to characterize part of the intensity of competition. As shown in the Table 7 in the appendix, a longer distance would result in a higher bid. At their respective average working distances, an additional 10 miles of distance would result in 0.27\% and 0.23\% increases in bids for the fringe and regular firm respectively.

\textsuperscript{18}We could not design a formal nonparametric test for the hypothesis that the fringe and regular firms are symmetric by using the frequency of winning because the number of bidders for each type is not fixed across auctions.

\textsuperscript{19}In the data we can only locate the county of a firm or a project. Then we use the distance between the central points of the two counties as the approximate distance between the a firm and a project. Given the relatively small size of a county, this approximated distance is acceptable.
example, for a project of approximately $1 million, 10 additional miles lead to less than a $3000 increase in bids, which is not significant relative to the project size. The results in Table 7 are also consistent with the auction theory: when the number of bidder increases, the firms submit more competitive bids. The Probit estimation results in Table 7 show that a longer distance decreases a firm’s probability to participate. However, when a firm has an unfinished project in the county where the new project is located, the firm is more likely to participate and submit more competitive bid. One explanation for this is that firm could potentially benefit from an economy of scale when several projects are conducted in the same county.

The reduced-form analysis above could give us a basic idea of firms’ bidding behavior in the procurement market. It cannot, however, give us much information on the cost structures of the bidding firms, which we can use to evaluate the efficiency of the procurement mechanism or to design a more efficient mechanism. In the following subsections we will recover the bidding firms’ unobserved project costs and their corresponding distributions in a structural way. Following the analysis above we will assume that the asymmetric IPC paradigm is applicable in the Ohio highway procurement market.

2.6.3 Empirical Results for Ohio Highway Procurement Auctions

We apply the conditional hazard rate based two-step nonparametric estimation method with its bandwidth selection rules to recover the distribution of project cost for each type of the firm. Since projects of extraordinarily large size occur sporadically in the data, we decide to choose projects that have an engineer’s estimate less than $10 million. 1231 projects fall into this group and account for 98% of all projects.
in the original data. The estimation in the first step is conditioned on any possible combination of \( N_1 \) and \( N_2 \), the number of fringe and regular firms. To obtain a good nonparametric estimate of the conditional density function, we only choose those combinations of \( N_1 \) and \( N_2 \) such that the number of observed bids for each type of firms in this subgroup is larger than 100.

From the first step we can estimate the unobserved project cost corresponding to each observed bid. After trimming, 93.20% data of regular firms and 92.28% data of fringe firms remain, which are considered as the unbiased estimates of project costs. We can then calculate the markup of all remaining bids, which is defined as \((b - c)/c\) reflecting the percentage increase on the real cost when a firm submits its bid. We report the results in Table 8. On average, the fringe firms have a markup of 10.72%, which is lower than 11.05% for the regular firms. This means the fringe firms bid more competitively than the regular firms on average. Consistent with the prediction by auction theory, both types of bidders bid more aggressively when the number of bidders increases. When there are 4 fringe (regular) firms in the auction, the average markup decreases to 6.75% (5.77%) for fringe (regular) firms.

From the second step we can estimate the conditional cost distribution functions of both types of firms for any given project size, i.e., the engineer’s estimate. We choose to report the result in the form of normalized cost, i.e., the estimated cost divided by the engineer’s estimate. The categorized histograms and density function estimates of the normalized cost for each type of firm are displayed in Figure 6 and Figure 7; the categorized summary statistics of the distribution of the normalized cost for each types of firms are reported in Table 9. From these results we can learn that on average the regular firms have a cost advantage over the fringe firms in this market.
in-depth categorization shows that for projects of relatively small size, the fringe firms actually enjoy a cost advantage. More specifically, we show the results of regression analysis between the estimated cost and the engineer’s estimate in Table 10. The firm-type dummy variable and the interaction term between this dummy variable and the engineer’s estimate have been included in the regression. The parameters for them are all statistically significant. The estimated parameters show that the engineer’s estimate-dependent cost function of regular firms has a higher intercept and lower slope and the engineer’s estimate-dependent cost function of fringe firms has a lower intercept and higher slope. By identifying the crossing point between these two cost functions, we can find that when the engineer’s estimate is less than about four hundred thousand dollars, the fringe firms on average have a lower cost than regular firms; when the engineer’s estimate is more than four hundred thousand dollars, the regular firms on average have a lower cost than fringe firms. This result is different from the usual point of view that in the highway procurement market the regular firms always dominate the fringe firms in terms of the construction cost. This observation could be explained as follows: the fringe firms may operate more flexibly in relatively small projects and thus enjoys some cost benefits; but when the project size is large, the regular firm enjoys the benefit from more efficient construction, more efficient management and economies of scales. Actually this unusual result is consistent with our observation in section 6.2. Given that frequency of bidding is controlled, the fringe firms have a higher probability of winning when the project size is relatively small and the regular firms have a higher probability of winning when the project size is relatively large. This empirical result also has a implication in policy design. A number of states in the U.S. have established programs to improve small or
fringe firms’ opportunities in the public procurement. The small or fringe firms are given preference treatment in the auction practice. This policy is more meaningful and justifiable when the project size is relatively large.

For the trimmed data, we can also do a counterfactual experiment: if the projects in the remaining data were procured through a second low-price auction instead of the first low-price auction, what would be the social cost to accomplish these projects? Auction theory tells us that the second low-price auction is efficient in the sense that the project would be offered to the firm with the lowest cost because each firm would bid their real costs. This is true even in the asymmetric setting where the first low-price auction is not always efficient. The total cost in our remaining sample is $633.097 million in the first low-price auction; the two-step estimation results show that the total cost would be $617.815 million in the second low-price auction. This means that the government can at least decrease the social cost by 2.41% through the second low-price auctions under the assumption that the second low-price auction mechanism would not discourage the participation of current bidding firms.

2.7 Conclusion

This chapter studies the empirically implementable bandwidth selection method for the structural-form nonparametric estimation of procurement auctions with observed heterogeneity in project sizes. When there exists heterogeneity of project size as in any typical auction data such that modeling the distribution of bidders’ project costs by using a conditional density function is more appropriate, the proposed bandwidth selection method produces much improved estimates of the unobserved costs and the distribution function of the costs than the commonly adopted one in the
literature and thus provides better policy guidance both in auction designs and in choosing a parametric functional form if a parametric estimation method is preferred. We have derived explicit expressions for the bandwidth selection rules that can be used for empirical study. These rules can also be easily adapted to the estimation in other auction mechanisms such as first-price sealed-bid auctions.

In addition to use the reference distribution method to derive the bandwidth selection rules, simulation based bandwidth selection method can also be used. The kernel smoothing method used in the two-step nonparametric estimators in this paper is subject to the boundary effects. The local polynomial smoothing method can be considered to reduce this bias.
CHAPTER 3

IDENTIFICATION AND ESTIMATION OF PROCUREMENT AUCTIONS WITH ENDOGENOUS ENTRY

3.1 Introduction

In most empirical auction literature, two basic assumptions have been maintained: first, the number of bidding firms in a procurement auction is publicly known to all firms before they submit their bids; second, the number of bidding firms is exogenous, in other words, it does not depend on the underlying project cost distribution. These assumptions have made the analysis of bidding strategy easier and the estimation of the underlying project cost distribution more tractable. Among others, one of the most important steps in analyzing an auction market is to test the relationship between the bid and the number of bidding firms. The test is aimed for two purposes: by a negative and statistically significant relationship between these two variables, on the one hand, it verifies the implication of the theoretical auction model; on the other hand, it justifies using the number of bidding firms in the bidding function to characterize the intensity of the competition. However, from both the theoretical and the empirical perspectives, there are settings in which these assumptions are not
satisfied. In this chapter, we study the procurement auction models with entry cost\textsuperscript{20}. In these models, a potential firm and a bidding firm need to be differentiated and a potential firm’s participation strategy needs to be considered.

Two different procurement auction models will follow, depending on when the entry cost is incurred. In the first model, the entry cost is incurred after a potential firm observes its project cost and thus can be explained as the bid-preparation cost. In this circumstance firms learn their private project costs before they decide whether to incur the bid-preparation cost. The bid-preparation cost can be explained as the cost of preparing the private bid. It can also be explained as the cost of travelling to the auction site or the opportunity cost of bidding. This entry cost exists especially in many procurement markets. When a project is procured, a detailed blueprint such as an architectural plan or a technical drawing and a detailed explanation of the item-by-item expenses are usually required upon submission of the bid. The nontrivial cost of preparing these documents can be regarded as the entry cost that the bidding firms need to spend. Samuelson (1985) provided an indirect analysis of the procurement auction with the bid-preparation cost by concentrating on the efficient mechanism design and Menezes and Monterior (2000) presented an direct analysis of the first-price sealed-bid auction model with the bid-preparation cost. To the best of our knowledge, no empirical analysis has been done in the auction setting with the bid-preparation cost.

In the second model, the entry cost is incurred before a potential firm observes its project cost and thus can be explained as the information-gathering cost. Potential

\textsuperscript{20}In this chapter, two different costs are introduced: the project cost and the entry cost. A firm needs to spend the project cost if and only if it is the winner in the procurement auction. A firm needs to spend the entry cost once it participates in the auction.
firms need to spend the information-gathering cost to estimate their private project costs. The classical model with the information-gathering cost is described as a two-stage game: at the first stage, each potential firm decides whether to incur the information-gathering cost before it learns its private project cost; at the second stage, the number of bidding firms is revealed and the participating firms submit their bids. Levin and Smith (1994) provided a theoretical analysis of bidders’ strategies in this setting. In equilibrium, firms will participate in the auction with some probability, which is endogenously determined by the underlying project cost distribution and the information-gathering cost. Li (2004) proposed a simulated GMM method to estimate a parametric first-price two-stage auction model with both the information-gathering cost and the effective reservation price. In addition, Li and Zheng (2005) proposed the structural inference for the procurement auction with the information-gathering cost. They extended the two-stage model to the case that the number of bidding firms is not revealed at the second stage, i.e., all potential firms are confronted with uncertain number of bidding firms.

In the following sections, we will study the identification and estimation of each auction model with the entry cost from the econometric perspective. We also study the identification in a broader sense: how to differentiate these two auction models from the observed data. Since the equilibrium participation and bidding strategies in these two models are fundamentally different, the differentiation between these two auction models is empirically important. We will maintain two assumptions for the analysis below. First, the number of bidding firms is not revealed to all potential firms, a more realistic assumption for the empirical analysis. Second, there exists a maximum level of bid $\bar{b}$ that a firm will submit, which belongs to the support of the
project cost\(^{21}\). The organization of this chapter is as follows. Section 2 presents the auction model with the bid-preparation cost, its implications and its nonparametric identification. Section 3 presents the auction model with the information-gathering cost, its implications and its nonparametric results. Section 4 studies the differentiation between the two procurement auction models with entry cost from both the economic and the econometric perspectives and the Bayesian structural estimation method. In section 5, we use Bayesian method to empirically analyze the Michigan highway procurement market. Section 6 concludes.

3.2 The Procurement Auction Model with Bid Preparation Cost

3.2.1 The Model Description and Implications

A single and indivisible project is auctioned to \(N\) potential firms through a low-price sealed-bid auction. All potential firms are risk-neutral and their respective private costs of completing the project are assumed to be identically and independently drawn from the distribution \(F(\cdot)\) with support \([c, \bar{c}]\). The procurement auction can be described as a two-stage game. At the first stage, each potential firm learns its private project cost and becomes an active firm if and only if its project cost is less than or equal to \(\tilde{b}\). At the second stage, each active firm decides whether to participate in the auction by incurring a participation cost \(k\) before submitting its bid, without knowing the number of bidding firms. The participation cost \(k\) can be

\(^{21}\)One example of \(\tilde{b}\) is the binding reservation price. The existance of \(\tilde{b}\) is necessary in the procurement auction model with entry. Without the existence of such \(\tilde{b}\), if a firm knows that it is the only bidder in the auction (when the number of actual bidders is revealed) or there is a positive probability that it is the only bidder in the auction (when the number of actual bidders is not revealed), the firm’s best strategy is to bid infinity or very high, which gives the bidder an infinite or high payoff or expected payoff. This conclusion is inconsistent with the observation in reality.
explained as the cost of preparing a bid or the opportunity cost of submitting a bid. If \( k \) is larger than \( \bar{b} - c \), then no firm will participate in the auction because the expected profit from submitting a bid cannot exceed \( \bar{b} - c \). In the following analysis we focus on the situation that \( k \in (0, \bar{b} - c) \). Then the probability that there exists at least one bidding firm is positive.

In this model, each potential firm needs to decide whether to participate in the auction and how much to bid if it participates. The following symmetric participation and bidding strategies are assumed: there exists a cutoff value \( c^* \in [\underline{c}, \bar{c}] \) such that an active firm \( i \) participates in the auction if and only if \( c_i \leq c^* \); if firm \( i \) participates, it bids according to the bidding function \( s(\cdot|F, N, c^*) \), which in general depends on the underlying project cost distribution, the number of potential firms and the cutoff value. As usual we will consider the symmetric, increasing and differentiable BNE bidding strategy. The following proposition characterizes the participation and bidding strategies.

**Proposition 4** In an independent private project cost low-price sealed-bid auction model, if there are \( N \geq 2 \) potential firms, the project cost of each firm is independently drawn from the distribution \( F(\cdot) \) with support \([\underline{c}, \bar{c}]\), the participation cost \( k \) belongs to the interval \((0, \bar{b} - c)\) where \( \bar{b} \) is the maximum possible bid a bidding firm will submit, then

(1) A potential firm participates if and only if its private project cost is less than or equal to a cutoff value \( c^* \) where \( c^* \) is uniquely determined by the following equation

\[
(\bar{b} - c^*)[1 - F(c^*)]^{N-1} = k
\]  

(3.1)
(2) If a potential firm participates, its bidding strategy is uniquely determined by

\[
s(c|F, N, c^*) = c + \frac{\int_{c}^{\pi^*} [1 - F(t)] N^{-1} dt + (\bar{b} - c^*)[1 - F(c^*)]N^{-1}}{[1 - F(c)]N^{-1}}
\]

(\text{3.2})

\[
= c + \frac{\int_{c}^{\pi^*} [1 - F(t)] N^{-1} dt + k}{[1 - F(c)]N^{-1}}
\]

for \(c \leq c^*\).

From the equilibrium participation and bidding strategies characterized by the equations (3.1) and (3.2), we can easily obtain the implications of the model. We summarize the property of the model in the following proposition:

**Proposition 5** (1) \(\partial c^*/\partial k < 0\), i.e., as the bid-preparation cost \(k\) increases, the cutoff value \(c^*\) decreases, ceteris paribus. It follows that the equilibrium participation probability of a potential bidder also decreases. (2) \(\partial c^*/\partial N < 0\), i.e., as the number of potential bidders \(N\) increases, the cutoff value \(c^*\) decreases, ceteris paribus. It follows that the equilibrium participation probability of a potential bidder also decreases. (3) At a given relevant project cost \(c\), the bidding function \(s(c|F, N, c^*)\) is not a monotone function of the number of potential bidders \(N\). (4) The expected contract cost does not necessarily decrease when the number of potential bidders increases.

The properties (1) and (2) show how the bid-preparation cost and the number of potential bidders can affect the cutoff value and the participation probability of a potential bidder. When the bid-preparation cost increases, only those bidders with more project cost advantage can afford to participate in the auction. So the cutoff value and the equilibrium participation probability will decrease. When the number of potential bidders increases, the competition from rivals increases, which would decrease the probability of winning for a potential bidder. So similarly, only those
bidders with more project cost advantage can afford to participate in the auction and both the cutoff value and the equilibrium participation probability will decrease.

The properties (3) and (4) present the results different from that in the fixed-\(n\) literature in which the number of bidding firms is known and exogenous so that bidding function and the expected contract cost are decreasing in the number of bidders. A detailed analysis of the bidding function in Proposition 5 shows that an increase in the number of potential bidders can have two effects on the bid: the competition effect and the entry cost effect. The competition effect means that more bidders with more project cost advantage would potentially participate in the auction and the probability of winning would decrease. So a bidding firm has to submit a more competitive bid, i.e., a lower bid. However, a bidding firm has to recover the nontrivial and irrevocable entry cost that is paid before submitting the bid. A low bid with the decreased probability of winning will end up with a negative expected profit. So a bidder has a tendency to increase the bid, which is considered as the entry cost effect. More specifically, it is easy to check the effect of an increase in the number of potential bidders \(N\) on the bidding function \(s(c|F, N, c^*)\). From the participation equation we have\(^{22}\)

\[
\frac{\partial c^*}{\partial N} = \frac{k \ln(1 - F(c^*))}{[1 - F(c^*)]^{N-1} + (\bar{b} - c^*)(N - 1)[1 - F(c^*)]^{N-2} f(c^*)} < 0
\]

From the bidding function and the equation above, we can get

\(^{22}\)For simplicity, we treat \(N\) as a continuous variable and take derivative w.r.t to it.
\[
\frac{db}{dN} = \frac{\partial s(c|F,N,c^*)}{\partial N} + \frac{\partial s(c|F,N,c^*)}{\partial c^*} \frac{\partial c^*}{\partial N} \\
= \int_c^{c^*} \left[ \frac{1 - F(t)^N}{1 - F(c)} \right]^{N-1} \ln \left[ \frac{1 - F(t)}{1 - F(c)} \right] dt \\
+ k \left[ \frac{1}{1 - F(c)} \right]^{N-1} \ln \left[ \frac{1}{1 - F(c)} \right] \\
+ \frac{[1 - F(c^*)]^{N-1}}{[1 - F(c)]^{N-1}} \frac{[1 - F(c^*)]}{[1 - F(c^*)]} k \ln(1 - F(c^*)) \\
+ (N - 1)kf(c^*)
\]

The first and third terms are negative, representing the competition effect; the second term is positive, representing the entry cost effect. At a high level project cost, the entry cost effect is dominating the competition effect, the bid will increase in the number of potential bidders; at a low level project cost, the competition effect is dominating the entry cost effect, the bid will decrease in the number of potential bidders. An example will be given below to show these effects.

The property (4) shows that the expected contract cost can increase in the number of potential bidders. It is well-known that in the fixed-\(n\) auction model without the endogenous participation, the procurer’s expected contract cost will always decrease when the number of bidders increases. The underlying reason is the exclusive competition effect. So in the fixed-\(n\) auction model, the procurer will expect more bidders to appear at the auction site. However, in the auction model with the endogenous entry such as that in Li and Zheng (2005) and the model considered here, more potential bidders do not necessarily result in lower expected contract cost. So whether the procurer should make more advertisements to attract more potential bidders is an empirical question that depends on the specific auction environment.

In the auction models with the endogenous entry, there is a positive possibility that no bidder will participate in the auction. Assume the project cost for the procurer
is $p_0$. Then procurer’s expected procurement cost conditional on the total number of potential bidders can be written as

$$E[\text{Cost}|N] = \int_{c^*}^{c} s(t|F, N, c) f_{c_{N, \text{min}}}(t) dt + \int_{c^*}^{\bar{c}} p_0 f_{c_{N, \text{min}}}(t) dt$$

where $s(\cdot|F, N, c^*)$ is the equilibrium bidding strategy and $f_{c_{N, \text{min}}}(\cdot)$ is the density function of the minimum of $N$ project costs independently and identically drawn from the density function $f(\cdot)$ on support $[c, \bar{c}]$, i.e.,

$$f_{c_{N, \text{min}}}(c) = N[1 - F(c)]^{N-1} f(c)$$

for $c \in [c, \bar{c}]$. Then it follows

$$E[\text{Cost}|N] = Nk F(c^*) + N \int_{c}^{c^*} [1 - F(t)]^{N-1} [tf(t) + F(t)] dt + p_0 [1 - F(c^*)]^N$$

In general we cannot determine the sign of $dE[\text{Cost}|N]/dN$. An increase in the number of potential bidders does not have a monotonic effect on the expected contract cost. The example below shows that more potential bidders can possibly increase the expected contract cost. In this circumstance, making more advertisements is not a good policy for the procurer.

Although we cannot observe the bid-preparation cost and test the property (2) empirically, the properties (3) and (4) do provide us some relationships between the observed variables that we can verify. If we observe the number of potential firms, the number of actual bidders and their respective bids in the data, then we can use the ratio of the number of bidding firms and the number of potential bidders as a proxy for the equilibrium participation probability and test how the number
of potential bidders affect the participation probability and the bid. Property (5) provides us with the possibility to evaluate the procurement auction mechanism after we estimate the structural model.

To illustrate the effect of an increase in the number of potential bidders on the cut-off value, the bidding function and the expected contract cost, we consider the following specific example. The project cost is drawn from the distribution $F(c) = 1 - (1 - c)^a$ for $c \in [0, 1]$ where $a = 4$. The participation cost $k = 0.3$. The procurer’s cost $p_0 = 1$. Conditional on the number of potential bidders, we first solve the participation cutoff value by applying the Newton method on the participation equation. Then we calculate the bidding function and the expected contract cost. In Figure 8 we draw the project cost density function, project cost distribution function, bidding functions and expected procurement contract cost corresponding to different number of potential bidders. In the example the expected contract cost is minimized at $N = 4$, implying that an increase in the number of potential bidders can possibly increase the expected procurement cost.

### 3.2.2 Nonparametric Identification of Auction Model with Bid Preparation Cost

In this section we study the identification of the auction model with the bid-preparation cost under the assumptions that the bid-preparation cost and the number of potential bidders are constant across repeated auctions. Given the distribution of the project cost, the participation strategy equation and the bidding function, the distribution of the equilibrium bid is uniquely determined. Let $G^*(\cdot)$ denote the distribution of bid on the support $[b, \bar{b}]$ and $g^*(\cdot)$ denote the corresponding density function. Note that $[b, \bar{b}] = [s(c|F, N, c^*), \bar{b}]$. It follows that for $\forall b \in [b, \bar{b}]$, 

\[ G^*(b) = \Pr(s(c) \leq b | c \leq c^*) \]
\[ = \Pr(c \leq s^{-1}(b) | c \leq c^*) \]
\[ = \frac{F(s^{-1}(b))}{F(c^*)} \]

and

\[ g^*(b) = \frac{f(s^{-1}(b))}{F(c^*)} \frac{1}{s'(s^{-1}(b))} \]

Then from the two equations above,

\[ \frac{f(s^{-1}(b))}{1 - F(s^{-1}(b))} \frac{1}{s'(s^{-1}(b))} = \frac{F(c^*) g^*(b)}{1 - F(c^*) G^*(b)} \]

By substituting the equation above into the first-order condition for the bidding function, we have

\[ c = b - \frac{1}{(N - 1) \frac{f(s^{-1}(b))}{1 - F(s^{-1}(b))} \frac{1}{s'(s^{-1}(b))}} \]
\[ = b - \frac{1}{(N - 1) \frac{g^*(b) F(c^*)}{1 - G^*(b) F(c^*)}} \]
\[ = b - \frac{1 - F(c^*) G^*(b)}{(N - 1) F(c^*) g^*(b)} \]
\[ = \xi(b, G^*, N, F(c^*)) \]

where \( \xi(\cdot, G^*, N, F(c^*)) \) is the inverse bidding function, characterized by the bid distribution \( G^*(\cdot) \), the number of potential bidders \( N \) and the equilibrium participation probability \( F(c^*) \). The proposition below shows that the conditions under which we can uniquely recover the structural elements of the model \( N, k \) and \( F(\cdot) \) given the bid distribution \( G^*(\cdot) \) and data \( \{n_l, b_{i,l}; i = 1, \cdots, n_l\}_{l=1}^L \), where \( n_l \) is the number of bidding firms in auction \( l \) for \( l = 1, \cdots, L \).
Proposition 6  Let $G^*(\cdot) \in P^{23}$ on $[b, \bar{b}]$. Let $\pi(\cdot)$ be a discrete distribution. There exists a distribution of firms’ project cost $F(\cdot) \in P$ , the number of potential firms $N \geq 2$ and the bid-preparation cost $k$ such that (1) $G^*(\cdot)$ is the equilibrium bid distribution in a low-price sealed-bid auction with the bid-preparation cost $k \in (0, \bar{b} - c)$ and (2) $\pi(\cdot)$ is the distribution of the number of bidding firms if and only if the following conditions hold:

C1: $\pi(\cdot)$ is binomial distribution with parameters $[N, F(c^*)]$ where $0 < F(c^*) < 1$.

C2: The observed bids $\{b_i\}_{i=1}^n$ are i.i.d. according to $G^*(\cdot)$ conditional on $n$.

C3: The function $\xi(\cdot, G^*, N, F(c^*))$ is strictly increasing on $[b, \bar{b}]$ and its inverse is differentiable on $[\xi(b, G^*, N, F(c^*)), \xi(\bar{b}, G^*, N, F(c^*))]$.

Moreover, if C1-C3 hold, then $N$, $k$ and $F(c^*)$ are unique and $F(\cdot)$ is uniquely defined on $[c, \xi(\bar{b}, G^*, N, F(c^*))]$ as $F(\cdot) = F(c^*)G^*(\xi^{-1}\cdot, G^*, N, F(c^*))$. In addition, $\xi(\cdot, G^*, N, F(c^*))$ is the quasi inverse of the equilibrium strategy $s(\cdot|F, N, c^*)$ in the sense that $\xi(b, G^*, N, F(c^*)) = s^{-1}(b|F, N, c^*)$ for all $b \in [b, \bar{b}]$.

Proposition 6 extends the identification result in Guerre et al. (2000) to the model with the bid-preparation cost. It shows that even the endogenous entry results from an unknown entry cost rather than the binding reservation price, the model can still be identified. However, the nonparametric estimation method proposed by Guerre et al can not be easily adapted to this model. Although $\xi(\cdot, G^*, N, F(c^*))$ is uniquely determined on $[b, \bar{b}]$, it can not be unbiasedly estimated at the boundary point $\bar{b}$ by the kernel-based nonparametric method. However, the definition of the bid-preparation cost in establishing the identification involves the value of $\xi(\bar{b}, G^*, N, F(c^*))$. More generally, under the circumstances that the bid-preparation cost and/or the number

$^{23}P$ denotes the set of absolute continuous distribution functions.
of potential bidders are heterogeneous across different auctions, the nonparametric identification and estimation will become too complicated to establish because of the interlinked nonlinear participation equation and nonlinear bidding function.

3.3 The Procurement Auction Model with Information Gathering Cost

3.3.1 The Model Description and Implications

A single and indivisible project is auctioned to $N$ potential firms through a low-price sealed-bid auction. The potential firms are risk-neutral and their respective private project costs of completing the projects are assumed to be identically and independently drawn from the distribution $F(\cdot)$ with support $[c, \bar{c}]$. The auction can be described as a two-stage game. At the first stage, each potential firm decides whether to spend the information-gathering cost $k$ and becomes an active bidder. At the second stage, each active firm observes its private project cost and submits a bid if the project cost is less than $\bar{b}$, without knowing the number of total actual bidders. The participation cost $k$ can be explained as the cost of estimating the private project cost. Let $\pi(b, c|q)$ denote the expected payoff of the bidding firm who optimally bids $b$ using a BNE strategy $s(\cdot)$ given its own project cost $c \leq \bar{b}$ and the symmetric entry probability $q$. Then the ex ante expected payoff for a potential bidder before observing the private project cost $c$ is

$$\int_{\xi}^{\bar{b}} \pi(b, c|q) f(c) dc$$
A potential bidder will participate in the procurement auction only when the ex ante expected payoff can cover the entry cost. The expected profit conditional on the entry probability $q$ can be written as

$$
\pi(b, c|q) = \sum_{i=0}^{N-1} \Pr(n = i)(b - c)[1 - F(s^{-1}(b))]^i
$$

$$
= (b - c) \sum_{i=0}^{N-1} C^i_{N-1} q^i (1 - q)^{N-1-i} \left[1 - F(s^{-1}(b))\right]^i
$$

$$
= (b - c) \left[q \left(1 - F(s^{-1}(b))\right) + 1 - q\right]^{N-1}
$$

$$
= (b - c) \left[1 - qF(s^{-1}(b))\right]^{N-1}
$$

Maximization of the expected payoff results in the following equation

$$
b = c + \left[1 - qF(c)\right]s'(c) \quad \text{(3.3)}
$$

Equation (3.3) and the boundary condition

$$
s(b) = \bar{b}
$$

characterize the BNE bidding function, which depends on the project cost distribution function $F(\cdot)$, the number of potential firms $N$ and the participation probability $q$:

$$
b = s(c|F, N, q) = c + \int_c^{\bar{b}} \left[1 - qF(t)\right]^{N-1} dt \quad \frac{1 - qF(c)}{(N - 1)qf(c)}
$$

for $c \leq \bar{b}$. By substituting the equilibrium bidding function $s(c|F, N, q)$ back to $\pi(b, c|q)$, we have

$$
\pi(b, c|q) = \int_c^{\bar{b}} [1 - qF(t)]^{N-1} dt
$$
for \( c \leq \bar{b} \). Then we can have the following proposition characterizing the equilibrium participation and bidding strategies in the auction model with the information-gathering cost.

**Proposition 7** In an independent private project cost low-price sealed-bid auction model, if there are \( N \geq 2 \) potential firms, the project cost of each firm is independently drawn from the distribution \( F(\cdot) \) with support \([c, \bar{c}]\), \( k \) is the information-gathering cost and \( \bar{b} \) is the maximum possible bid a bidding firm will submit, then

1. A potential firm will incur the information-gathering cost and participate in the auction with probability \( q^* \). \( q^* \) satisfies

\[
q^* = \begin{cases} 
0, & \text{if } k \geq \int_{c}^{\bar{b}} F(t) \, dt \\
1, & \text{if } k \leq \int_{c}^{\bar{b}} [1 - F(t)]^{N-1} F(t) \, dt \\
q, & \text{otherwise}
\end{cases}
\]

where \( q \) is the unique solution to \( \int_{c}^{\bar{b}} [1 - q F(t)]^{N-1} F(t) \, dt = k \).

2. For a bidding firm, the equilibrium bidding function is

\[
b = s(c|F, N, q^*) = c + \frac{\int_{c}^{\bar{b}} [1 - q^* F(t)]^{N-1} dt}{[1 - q^* F(c)]^{N-1}}
\]

The corresponding properties of the auction model with the information-gathering cost is summarized in the following proposition:

**Proposition 8** (1) \( \frac{\partial q^*}{\partial k} \leq 0 \), i.e., as the information-gathering cost \( k \) increases, the equilibrium participation probability \( q^* \) decreases. (2) \( \frac{\partial q^*}{\partial N} \leq 0 \), i.e., as the number of potential firms \( N \) increases, the equilibrium participation probability \( q^* \) decreases. (3) At a given relevant private project cost \( c \), the bidding function \( s(c|F, N, q^*) \) is not necessarily a monotone function of the number of potential bidders \( N \).
Properties (1) and (2) are similar to that in the auction model with the bid-preparation cost and are quite intuitive. To understand property (3), note that

\[ \frac{db}{dN} = \frac{\partial s(c|F, N, q^*)}{\partial N} + \frac{\partial s(c|F, N, q^*)}{\partial q^*} \frac{\partial q^*}{\partial N} \]

Form the equilibrium bidding function, we have

\[ \frac{\partial s(c|F, N, q^*)}{\partial N} = \int_c^{\bar{q}} \left[ \frac{1 - q^*F(t)}{1 - q^*F(c)} \right]^{N-1} \ln \left[ \frac{1 - q^*F(t)}{1 - q^*F(c)} \right] dt < 0 \]

and

\[ \frac{\partial s(c|F, N, q^*)}{\partial q^*} = (N-1) \int_c^{\bar{b}} \left[ \frac{1 - q^*F(t)}{1 - q^*F(c)} \right]^{N-2} \frac{F(c) - F(t)}{[1 - q^*F(c)]^2} dt < 0 \]

From the equilibrium participation strategy, given \( q^* \in (0, 1) \), we have

\[ \frac{\partial q^*}{\partial N} = \frac{\int_{\bar{q}}^\bar{b} \left[ 1 - q^*F(t) \right]^{N-1} F(t) \ln[1 - q^*F(t)]dt}{(N-1) \int_{\bar{q}}^\bar{b} \left[ 1 - q^*F(t) \right]^{N-2} F^2(t) dt} < 0 \]

Then it follows that the first term \( \frac{\partial s(c|F, N, q^*)}{\partial N} \) above is negative, indicating a competition effect; the second term \( \frac{\partial s(c|F, N, q^*)}{\partial q^*} \frac{\partial q^*}{\partial N} \) above is positive, indicating an entry effect\(^{24}\), so \( \frac{db}{dN} \) is not a monotone function of \( N \). It follows that \( b \) is not necessarily decreasing in \( N \). We can illustrate the effect of the number of potential bidders on the bidding function using a specific example. The project cost is drawn from the normal distribution truncated on \([0, 6]\) with mean 3 and standard deviation 1. Suppose the participation cost \( k = 1 \) and \( \bar{b} = 6 \). In Figure 9 we draw bidding functions corresponding to number of potential bidders \( N = 3 \) and \( N = 8 \), which presents a result different from fixed-\( n \) auction model.

\(^{24}\) Li and Zheng (2005) identified the same compounding competition and entry effect in their model.
3.3.2 Nonparametric Identification of Auction Model with Information Gathering Cost

In the model with the information-gathering cost, if \( \bar{b} < \bar{c} \), then the number of bidding firms is endogenously determined twice. The first results from the randomized participation strategy. The second results from the fact that firms will submit bids if and only if their project costs are less than \( \bar{b} \). Then we have the following proposition.

**Proposition 9** If \( \bar{b} < \bar{c} \), the procurement auction model with the information-gathering cost cannot be identified.

In the following we will focus on the case that \( \bar{b} = \bar{c} \), i.e., the endogenous participation only results from the information-gathering cost. As in the previous section, we can derive the inverse bidding function based on the distribution function of bid \( G(\cdot) \). By definition,

\[
G(b) = \Pr(s(c) \leq b) = F(s^{-1}(b))
\]

and

\[
g(b) = \frac{f(s^{-1}(b))}{s'(c)}
\]

Substituting \( G(\cdot) \) and \( g(\cdot) \) into equation (3.3), we have

\[
c = b - \frac{1 - q^*G(b)}{(N - 1)q^*g(b)} = \xi(b, G, N, q^*)
\]

which depends on the bid distribution function \( G(\cdot) \), the equilibrium participation probability \( q^* \) and the number of potential bidders \( N \). The following proposition
establishes the nonparametric identification result for the auction model with the information-gathering cost.

**Proposition 10** Let $G(\cdot) \in P^{25}$ on $[b, \overline{b}]$. Let $\pi(\cdot)$ be a discrete distribution. There exists a distribution of project cost $F(\cdot) \in P$, $N \geq 2$ potential firms and the information-gathering cost $k > 0$ such that (1) $G(\cdot)$ is the equilibrium bid distribution in a low-price sealed-bid auction with the formation-gathering cost $k$ and (2) $\pi(\cdot)$ is the distribution of the number of actual bidders if and only if the following conditions hold:

1. $\pi(\cdot)$ is binomial distribution with parameters $[N, q^*]$ where $0 < q^* < 1$.
2. The observed bids $\{b_i\}_{i=1}^{n}$ are i.i.d. as $G(\cdot)$ conditional on $n$ and $\lim_{b \uparrow \overline{b}} g(b) = +\infty$.
3. The function $\xi(\cdot, G, N, q^*)$ is strictly increasing on $[b, \overline{b}]$ and its inverse is differentiable on $[c, \overline{c}] = [\xi(b, G, N, q^*), \overline{b}]$.

Moreover, if C1-C3 hold, then $N$, $q^*$ and $k$ are unique and $F(\cdot)$ is uniquely defined on $[c, \overline{c}]$ as $F(\cdot) = G(\xi^{-1}(\cdot, G, N, q^*))$. In addition, $\xi(\cdot, G, N, q^*)$ is the quasi inverse of the equilibrium strategy $s(\cdot|F, N, q^*)$ in the sense that $\xi(b, G, N, q^*) = s^{-1}(b|F, N, q^*)$ for all $b \in [b, \overline{b}]$.

Proposition 10 above shows that the distribution of bid and the distribution of the number of bidders can be used to uniquely determine the underlying project cost distribution, the number of potential bidders and the information-gathering cost. However, as in the case of the bid-preparation cost, when the entry cost and/or the number of potential bidders are heterogeneous across different auctions, the nonparametric identification and estimation will become too complicated to establish because of the interlinked nonlinear participation equation and nonlinear bidding function.

$^{25}P$ denotes the set of absolute continuous distribution functions.
3.4 General Identification of Auction Models with Entry Cost and Structural Estimation

In this section, we assume that the endogenous entry only results from the existence of entry cost, i.e., $\tilde{b} = \tilde{c}$. The previous two sections present two competing auction models with the entry cost and their respective properties. In both models we make similar assumptions and derive similar implications that can be used to explain the observed data. For example, we assume that the number of bidding firms is not publicly known; both models can be used to model the firms’ endogenous participation; the number of bidding firms follows the binomial distribution; an increase in entry cost or number of potential firms has the similar effect on potential firms’ participation strategy. However, there also exists fundamental difference between these two models. In the model with the information-gathering cost, a potential firm’s participation strategy does not depend on its private project cost, i.e., any bidding firm’s project cost is not truncated. So the distribution of project cost can be identified over the full support $[\underline{c}, \overline{c}]$. In the model with the bid-preparation cost, only those bidders whose project costs are below some endogenously determined cut-off level $c^*$ will participate in the auction. So the distribution of the project cost can only be identified over the partial support $[\underline{c}, c^*]$. From the economic perspective, if it is reasonable to expect that a potential firm’s participation strategy depends on its private project cost, i.e., a potential firm with lower private project cost is more likely to participate in the auction than a potential firm with higher private project cost, in other words, the participation implies the project cost advantage, only the auction model with the bid-preparation cost is appropriate to rationalize this expectation.
In practice, we may only observe the potential firms’ participation strategies and the bids of bidding firms and do not have enough information about the underlying mechanism that generates the endogenous entry. The following proposition shows the possibility of identifying the auction model with entry from the data.

**Proposition 11** Assume that \( g^{\text{info}}(\cdot) \) and \( g^{\text{prep}}(\cdot) \) denote the density functions of the equilibrium bid in the auction model with the information-gathering cost and the bid-preparation cost. Then the auction model with the information-gathering cost implies that \( \lim_{b \to \infty} g^{\text{info}}(b) = +\infty \); while the auction model with the bid-preparation cost implies that \( \lim_{b \to \infty} g^{\text{prep}}(b) < \infty \). More generally, for auction data with heterogeneity in the project size and the number of potential firms, we have \( \lim_{b \to \infty} g^{\text{info}}(b|x, N) = +\infty \) and \( \lim_{b \to \infty} g^{\text{prep}}(b|x, N) < \infty \) where the variable \( x \) is used to characterize the observed heterogeneity of the project size and \( \bar{b}(x) \) is upper bound of bid in an auction of the size \( x \).

Propositions (6) and (10) show the respective restrictions that the procurement auction models with the bid-preparation cost and the information-gathering cost impose on the distributions of bid and the number of bidding firms. Proposition (11) identifies the difference between equilibrium bid distributions that can be used to differentiate the procurement auction model with the bid-preparation cost and the procurement auction model with the information-gathering cost. We can understand the results in Proposition (11) from the underlying assumptions in the procurement auction models with different entry cost. When the entry cost is incurred before bidders can observe their project cost, as in the model with the information-gathering cost, the entry cost is a sunk cost and has no effect on the bidding strategy given that a bidder has participated in the procurement auction. But when the entry cost
is incurred when or after bidders observe their project cost, as in the model with the bid-preparation cost, bidders will optimally choose the bids in order to recover the entry cost incurred, i.e., the entry cost has effect on the bidding strategy. So bidders will behave differently in these two models as shown above.

It is well recognized that the structural econometric auction model is non-regular in the sense that the support of the response variable depends on the distribution (or parameters of the distribution) of the latent variables. It is also known that the estimation of the structural econometric auction model usually involves tremendous numerical burdens. When the endogenous entry is considered and heterogeneity across different auctions exists in the data, the econometric model becomes more complicated. The nonparametric method and the usual likelihood or moment-based parametric estimation methods, if still feasible, are difficult to implement. The Bayesian approach is a preferred approach for inference in the complicated structural auction model. For example, Albano et al (1998) proposed a Bayesian approach to the econometrics of first-price auctions. Albano et al (1998) proposed Bayesian inference in repeated English auctions. Bajari (1998) used Bayesian method to study the econometrics of sealed-bid auctions. Li and Zheng (2006) used the semiparametric Bayesian method to estimate an auction model with the information-gathering cost.

Both auction models with entry cost can be analyzed by Bayesian method in a unified structure. The latent variables are the unobserved project cost $c$ and the entry cost $k$. Assume that $c$ follows the distribution $F(\cdot)$ on the support $[c_l, c_u]$ and $k$ follows the distribution $P(\cdot)$ on the support $[k_l, k_u]$. The response variables are the observed bid $b$ and the number of bidders $n$. The distribution function $F(\cdot)$, the number of potential bidders $N$ and a random realization of the entry cost $k$ uniquely determine
the equilibrium probability of participation, denoted by \( q^* \equiv q^*(F, N, k) \). Assume the BNE strategy is \( s(F, N, q^*) \). Then we can derive the density functions of the response variables:

\[
\pi(b|F, N, q^*) = \frac{f(\phi(b|F, N, q^*))}{F'(c^*)} \left| \frac{\partial \phi(b|F, N, q^*)}{\partial b} \right| \times \mathbf{1}(b \in [s(c|F, N, q^*), \bar{b}])
\]

and

\[
\pi(n|N, q^*, n \geq 1) = \frac{C_n^m(q^*)^n(1 - q^*)^{N-n}}{1 - (1 - q^*)^N}
\]

where \( \phi(b|F, N, q^*) \) is the inverse bidding function and \( c^* = \bar{b} \) for model with the information-gathering cost and \( c^* = F^{-1}(q^*) \) for model with the bid-preparation cost. Posterior distributions can be derived from these two density functions and Bayesian inference can be implemented. In the following section, we will use this method to empirically analyze the Michigan highway procurement auctions.

### 3.5 Empirical Analysis of Michigan Highway Procurement Auctions

#### 3.5.1 Market and Data Description

The Michigan Department of Transportation (MDOT) uses the low-price sealed-bid auction to contract out the highway construction and maintenance projects. We collect the data from the official website of MDOT\(^ {27} \). The data consists of 579 projects, which are paving or resurfacing projects in 2004 and 2005. We choose paving or resurfacing as the main work type because they most closely fit the independent private cost setting as shown by Hong and Shum (2002). The information we have in the data includes the letting date, the start date, the planned completion date, the inverse bidding function and \( c^* = \bar{b} \) for model with the information-gathering cost and \( c^* = F^{-1}(q^*) \) for model with the bid-preparation cost. Posterior distributions can be derived from these two density functions and Bayesian inference can be implemented. In the following section, we will use this method to empirically analyze the Michigan highway procurement auctions.

\(^{26}\)This density function results from the fact that we usually only observe the data for auctions with at least 1 actual bidding firm.

\(^{27}\)Please refer to the website http://mdotwas1.mdot.state.mi.us/public/bids/ for details.
the engineer’s estimate, the mileage of the project, the identity of all plan holders, the identity and bids of all bidding firms. The summary statistics of the data is presented in Table 11. We assume that all firms are homogeneous ex ante since we do not observe any asymmetry ex ante among the bidding firms.

We use the reduced-form estimation method to check the implications of our data. First we try to identify the relationship between the bid and the number of potential/bidding firms. In the data of Michigan highway procurement auctions, we observe the number of plan holders in addition to the number of bidding firms in each auction. Since only those firms who buy project plans from MDOT are eligible to bid in the auction, the plan holders in an auction, which are public information, can be regarded as the potential firms in that auction. We use regression analysis to detect whether the bidding firms know the total number of bidding firms or they only know the number of potential firms in deciding their bids. We regress the bids on the number of bidding firms and other covariates in the data for cases with and without auction dummy variables; we also regress the normalized bids (bid divided by the engineer’s estimate) on the number of bidding firms, other covariates and the auction dummy variables. The regression results are presented in Table 12. We do all the same regressions as above again except that the number of bidding firms are substituted by the number of potential firms. The regression results are presented in Table 13. We summarize all results in Table 14. In all the regressions, the number of bidding firms does not have a significant effect on the bid or normalized bid. So this is consistent with the model assumption that the firms do not know the number of bidding firms. On the contrary, the number of potential firms has significant effects on the bid and normalized bid. Furthermore, when the squared number of potential
firms is also included in the regression, we can find that bid and normalized bid is decreasing in the number of potential bidders when it is not too larger, and bid and normalized bid can increase in the number of potential bidders when it becomes large. This is consistent with the assumption and implication of the theoretical model: the firms only take the number of potential firms into account when making their bidding decisions and the bid is not always decreasing in the number of potential firms. As usual, the regression results show that the engineer’s estimate is the most important factor to explain the change in the observed bids. However, engineer’s estimate can only explain the variations of bids across auctions, it cannot explain the variations of bids within an auction. Second we try to identify the relationship between the equilibrium participation probability and the number of potential firms. We estimate the relationship between the number of bidding firms and log of the number of potential firms and other covariates by a Poisson model. We also use the ratio of the number of bidding firms and the number of potential firms as the proxy for the equilibrium participation probability and regress it on log of the number of potential bidders and other covariates. The results are presented in Table 15 and Table 16. Both results confirms that the participation probability is decreasing in the number of potential bidders, which is consistent with the model implication. Third, we display the distribution of the errors resulting from the regression of the bid on the auction-specific covariates in Figure 11. It is obvious that the distribution reaches its mode in the interior of the support and thus this is an evidence to support the auction model with the bid-preparation cost. From the analysis above, we conclude that the auction model with the bid-preparation cost is appropriate to explain the observed participation and bidding strategies in the highway procurement auctions.
3.5.2 Implementation of the Structural Estimation for Entry and Bidding

We will use the procurement auction model with the bid-preparation cost to empirically analyze the Michigan highway procurement auctions. We follow the parametric assumptions used in Li and Zheng (2006). More specifically, the project cost $c$ follows the Exponential distribution:

$$f(c|\mu) = \frac{1}{\exp(\mu)} \exp[-\frac{1}{\exp(\mu)}c]$$

for $c \in (0, +\infty)$. $\mu$ is parameterized as follows:

$$\mu = \alpha + x\beta + u$$

where $x$ is the covariate vector that characterizes the observed heterogeneity across different auctions and the random variable $u$ captures the unobserved heterogeneity in different auctions\textsuperscript{28}. By including the constant $\alpha$ in the equation, we can normalize $u$ by assuming that $E(u) = 0$. The corresponding distribution function of the project cost is

$$F(c|\mu) = 1 - \exp[-\frac{1}{\exp(\mu)}c]$$

The latent variables in each auction are $\mu$ and $c^*$, from which we can derive the equilibrium bidding function and the equilibrium probability of participation. However, in the Bayesian MCMC simulation, we will focus on $\mu$ and $q^* \equiv F(c^*|\mu)$ instead,

\textsuperscript{28}It has been shown in the reduced-form regressions that the inclusion of auction dummy variables can improve the $R^2$, which is an evidence to support unobserved heterogeneity.
where \( q^* \) is the equilibrium probability of participation. Note that there exists one-to-one correspondence between \( q^* \) and \( c^* \) conditional on \( \mu \), so it makes no essential difference to choose between \((\mu, c^*)\) and \((\mu, q^*)\). But the natural restriction on \( q^* \), i.e., \( q^* \in [0, 1] \), makes it easier to choose the jumping distribution in the Metropolis algorithm. Using the results in the section 2 and the project cost distribution assumption above, the bidding function conditional on \( \mu \) and \( q^* \) is as follows:

\[
\begin{align*}
\bar{b} &= s(c|\mu, q^*) \\
&= c + \int_c^\infty \left[ 1 - F(t|\mu) \right]^{N-1} dt + \left[ 1 - F(c^*|\mu) \right]^{N-1} (\bar{b} - c^*) \\
&= c + (1 - q^*)^{N-1} \exp\left( \frac{(N-1)c^*}{e^\mu} \right) \left[ \bar{b} + e^\mu \ln(1 - q^*) - \frac{e^\mu}{N - 1} \right] + \frac{e^\mu}{N - 1}
\end{align*}
\]

for \( c \leq c^* = -e^\mu \ln(1 - q^*) \). It follows that

\[
\begin{align*}
s'(c|\mu, q^*) &= \frac{\partial s(c|\mu, q^*)}{\partial c} \\
&= 1 + \frac{N - 1}{e^\mu} (1 - q^*)^{N-1} \exp\left( \frac{(N-1)c^*}{e^\mu} \right) \left[ \bar{b} + e^\mu \ln(1 - q^*) - \frac{e^\mu}{N - 1} \right]
\end{align*}
\]

Let \( c = \phi(b|\mu, q^*) \) denote the corresponding inverse bidding function conditional on \( \mu \) and \( q^* \). In the procurement auction with the bid-preparation cost, there is a positive probability that no bidder will participate in the auction. This happens either because that the entry cost is too high such that no firm will participate or all potential firms’ realized project cost are not low enough. It follows that we can observe an procurement auction in the data only under the following two conditions: first, the equilibrium participation probability is positive; second, the number of actual bidding firms is positive. We will take these two conditions into account in deriving the posterior density functions. From the equilibrium bidding function, we can derive the probability density function of the equilibrium bid,
\[ g(b|\mu, q^*) = \frac{1}{\exp(\mu)} \exp\left[-\frac{1}{\exp(\mu)} \phi(b|\mu, q^*)\right] \left| \frac{\partial \phi(b|\mu, q^*)}{\partial b} \right| \times 1 \left( b \in [b(\mu, q^*), \bar{b}] \right) \]

where

\[ \frac{\partial \phi(b|\mu, q^*)}{\partial b} = \frac{1}{s'(c|\mu, q^*)} \]

and

\[ \bar{b}(\mu, q^*) = (1 - q^*)^{N-1} \left[ \bar{b} + e^\mu \ln(1 - q^* - \frac{e^\mu}{N - 1}) + \frac{e^\mu}{N - 1} \right] \]

The entry cost \( k \) is assumed to follow the exponential distribution as follows:

\[ p(k|x, u) = \frac{1}{\exp(\gamma + x\delta + u)} \exp\left[-\frac{1}{\exp(\gamma + x\delta + u)} k\right] \]

for \( k \in [0, +\infty) \). As in Li and Zheng (2006), define \( \delta^* = \beta - \delta, \gamma^* = \alpha - \gamma, \theta = (\gamma^*, \delta^*) \) and \( x^* = (1, x) \). Then we can write \( p(k|x, u) \) as

\[ p(k|\mu, \theta) = \frac{1}{\exp(\mu - x^*\theta)} \exp\left[-\frac{1}{\exp(\mu - x^*\theta)} k\right] \]

The corresponding distribution function is

\[ P(k|\mu, \theta) = 1 - \exp\left[-\frac{1}{\exp(\mu - x^*\theta)} k\right] \]

From the equilibrium participation equation and the fact \( q^* > 0 \), we can derive the conditional density function of \( q^* \) as

\[ q(q^*|\mu, \theta, q^*) = p(k|\mu, \theta) \times \left| \frac{\partial k}{\partial q^*} \right| \frac{1}{P(b|\mu, \theta)} \times 1 \left( q^* \in (0, 1] \right) \]

\[ = \frac{1}{\exp(\mu - x^*\theta)} \exp\left[-\frac{1}{\exp(\mu - x^*\theta)} k\right] \times \left| \frac{\partial k}{\partial q^*} \right| \times 1 \left( q^* \in (0, 1] \right) \]
where

\[ k = [\bar{b} + e^\mu \ln(1 - q^*)](1 - q^*)^{N-1} \]

and

\[
\frac{\partial k}{\partial q^*} = -(N - 1)(1 - q^*)^{N-2}[\bar{b} + e^\mu \ln(1 - q^*)] - (1 - q^*)^{N-2} e^\mu
\]

The distribution of the unobserved heterogeneity \( u \) is left unspecified and will be estimated by Bayesian nonparametric density estimation using Dirichlet process prior. Assume \( d_l \) and \( \sigma^2_l \) are the mean and variance of the normal component in auction \( l \). We will estimate the parameters \( \alpha, \beta \) and \( \theta \) and the distribution of the error term \( u \) by the Bayesian method. We assume the following priors on the parameter distribution \( \beta \sim N(\beta_0, D_{\beta0}^{-1}) \) and \( \theta \sim N(\theta_0, D_{\theta0}^{-1}) \). The latent variables \( \{\mu_l, q^*_l\} \) for \( l = 1, \cdots, L \) will be augmented by the MCMC simulation, which are based on the blocks \( \{\mu_l, q^*_l\}_{l=1}^L, \beta, \theta, \{d_l, \sigma^2_l\}_{l=1}^L \) and the associated full conditional distributions of

\[
[\mu_l, q^*_l | \beta, \theta, d_l, \sigma^2_l, b_l, n_l]
\]

\[
[\beta | \mu, d, \sigma^2]
\]

\[
[\theta | \mu, q^*]
\]

More specifically, the Metropolis-within-Gibbs algorithm is based on the following posterior distributions. First we draw \( (\mu_l, q^*_l) \) for \( l = 1, \cdots, L \). The posterior joint distribution of \( (\mu_l, q^*_l) \) is
\[ \pi(\mu_i, q_i^* | \beta, \theta, d_i, \sigma^2_i, b_i, n_i) \propto \]
\[ \pi(\mu_i, q_i^* | \beta, \theta, d_i, \sigma^2_i, b_i, n_i) \propto \]
\[ \pi(b_i, n_i | \mu_i, q_i^*, \beta, \theta, d_i, \sigma^2_i) \pi(q_i^* | \mu_i, \beta, \theta, d_i, \sigma^2_i) \pi(\mu_i | \beta, \theta, d_i, \sigma^2_i) \propto \]
\[ \prod_{i=1}^{n_i} \frac{1}{\exp(\mu_i)} \exp\left[-\frac{1}{\exp(\mu_i)} \phi(b_{ii} | \mu_i, q_i^*) \right] \left| \frac{\partial \phi(b_{ii} | \mu_i, c_i^*)}{\partial b} \right| \]
\[ \times \prod_{i=1}^{n_i} 1 \left( b_{ii} \in \left[ \bar{b}(\mu_i, q_i^*), \bar{b}(\mu_i, q_i^*) \right] \right) \times \frac{\left(N_i\right)(q_i^*)^{n_i}(1 - q_i^*)^{(N_i-n_i)}}{1 - (1 - q_i^*)^{N_i}} \]
\[ \times \frac{1}{\exp(\mu_i - x_i^*\theta)} \exp\left[-\frac{1}{\exp(\mu_i - x_i^*\theta)} k_i \right] \]
\[ \times \frac{1}{1 - \exp\left[-\frac{b_i}{\exp(\mu_i - x_i^*\theta)} \right]} \left| \frac{\partial k_i}{\partial q_i^*} \right| \times 1 (q_i^* \in (0, 1]) \]
\[ \times \frac{1}{\sqrt{2\pi\sigma^2_i}} \exp\left[-\frac{(\mu_i - x_i)^2}{2\sigma^2_i} \right] \]

where

\[ \frac{\partial \phi(b_{ii} | \mu_i, c_i^*)}{\partial b} = \frac{1}{s'(c_{ii} | \mu_i, c_i^*)} \]

\[ k_i = (1 - q_i^*)^{N_i-1} \left[ \bar{b} + e^{\mu_i} \ln(1 - q_i^*) \right] \]

and

\[ \frac{\partial k_i}{\partial q_i^*} = - (N_i - 1)(1 - q_i^*)^{N_i-2} \left[ \bar{b} + e^{\mu_i} \ln(1 - q_i^*) \right] - (1 - q_i^*)^{N_i-2} e^{\mu_i} \]

Second, we draw \( \beta \). The posterior joint distribution of \( \beta \) is
Third, we draw \( \theta \). The posterior joint distribution of \( \theta \) is

\[
\begin{align*}
    f(\theta|\mu, q^*) &\propto \\
    f(\theta, \mu, q^*) &\propto \\
    f(q^*|\theta, \mu) &\propto \\
    \exp[-(\theta - \theta_0)'D_0(\theta - \theta_0)/2]
\end{align*}
\]

\[
\times \prod_{i=1}^{L} \frac{1}{1 - \exp[-\frac{\beta}{\exp(\mu_i - x_i^*\theta)}]} \frac{1}{\exp(\mu_i - x_i^*\theta)} \exp\left[-\frac{1}{\exp(\mu_i - x_i^*\theta)} k_i \right] \\
\times \left| \frac{\partial k_i}{\partial q_i^*} \right| \times 1 \quad (q_i^* \in (0, 1))
\]

where

\[
k_i = (1 - q_i^*)^{N_i - 1} \left[ \bar{b} + e^{\mu_i} \ln(1 - q_i^*) \right]
\]

and

\[
\frac{\partial k_i}{\partial q_i^*} = -(N_i - 1)(1 - q_i^*)^{N_i - 2} \left[ \bar{b} + e^{\mu_i} \ln(1 - q_i^*) \right] - (1 - q_i^*)^{N_i - 2} e^{\mu_i}
\]

Last, we update \( d_l \) and \( \sigma_l^2 \) for \( l = 1, \cdots, L \). Detailed description of the Metropolis-within-Gibbs algorithm are given in the appendix.
3.5.3 Empirical Results for Michigan Highway Procurement Auctions

The variables we use in the Bayesian estimation include $\ln(Eng.Est.)$, $\ln(Mileage)$ and $\ln(Days)$. Given that the resurfacing projects are not complicated, these variables can control well the observed heterogeneity of different projects. Except that we believe $\ln(Eng.Est.)$ will have a large positive impact on the project cost distribution, we have little information about how the covariates will affect the distributions of project cost and the bid-preparation cost. So the vague prior distributions of the parameter vector $\beta$ and $\theta$ are chosen as follows: $\beta \sim N(0_3, 10 \times I_3)$ and $\theta \sim N(0_4, 10 \times I_4)$. We run the MCMC simulation 20000 times. As usual, the initial 5000 times are taken as the burn-in period and the last 15000 times are used to estimate the parameters of interest or run post-estimation simulation experiments. On the platform of Pentium 4 2GHz CPU and 1G RAM, it took about 36 hours to finish 20000 random draws for all latent variables and parameters. Much computation is involved because of the nonlinear bidding function.

In the Bayesian MCMC simulation, Metropolis sampling is used for the latent variables $(\mu_l, q^*_l)$ and the parameter vector $\theta$. The average acceptance rates for sampling $(\mu_l, q^*_l)$ is 24.7% and the acceptance rate for sampling $\theta$ is 76.8%. Figure 12 in the appendix shows the time series plots of $\mu_l$ and $q^*_l$ for a typical procurement auction in the data. Figure 13 displays the time series plot and the histogram for the Bayesian MCMC draws of the parameters in the distributions of project cost and the bid-preparation cost. We summarize the parameter estimates in Table 17. The mean and standard deviation for each parameter are calculated using 15000 MCMC draws after the initial 5000 burn-in period. From the estimates of the parameters in
the project cost distribution, we find that the variable $\ln(\text{Eng.Est.})$ has a large and significant effect on the mean of project cost distribution. This is consistent with the results in the literature on highway procurement auctions. The Engineer’s Estimate is the best single variable to characterize the project size and thus is highly correlated with the mean of the firms’ project cost. It is worth mention that the estimated coefficient of $\ln(\text{Eng.Est.})$ is 1.1787, which shows a larger effect of Engineer’s Estimate on mean project cost than what is usually reported in the literature, especially in Li and Zheng (2005). This difference is consistent with our model implication. The bidding firms in our model are those with the project cost advantage. But $\exp(\mu)$ in the model specification is the mean of project cost distribution of the whole potential firm population, which is larger than the mean of project cost distribution of the bidding firms. The variable $\ln(\text{Mileage})$ has a significant negative effect on the mean of the project cost distribution. This is consistent with the results in the reduced-form analysis we did: the bids tend to decrease with Mileage after the size of the project is controlled. The variable $\ln(\text{Days})$ has a nonsignificant effect on the mean of the project cost distribution. The estimates of the parameters in the bid-preparation cost distribution show that the variable $\ln(\text{Mileage})$ has a positive effect on the mean bid-preparation cost; on the contrary, the variables $\ln(\text{Eng.Est.})$ and $\ln(\text{Days})$ have negative effects on the bid-preparation cost. Without controlling the correlation between $\ln(\text{Eng.Est.})$ and $\ln(\text{Mileage})$, the bid-preparation cost tends to be larger in the procurement auction of larger size. One interesting result is that given the project size is controlled, the bid-preparation cost tends to be smaller in

\footnote{The parameter estimate for the variable $\log(\text{Eng. Est.})$ in Li and Zheng (2005) is 0.9425. The most important reason for this difference is the endogenous selection of bidding firms with project cost advantage in the model with bid-preparation cost.}
the procurement auction of longer length. So if the government extends the working length of a project, it is possible that the decrease in the bid-preparation cost can reduce the expected procurement cost. Note that the effect of the decrease in the bid-preparation cost on the bid can be written as

\[
\frac{db}{dk} = \frac{\partial s(c|F, N, c^*)}{\partial k} + \frac{\partial s(c|F, N, c^*)}{\partial c^*} \frac{\partial c^*}{\partial k}
\]

Since

\[
k = (\bar{b} - c^*)[1 - F(c^*)]^{N-1}
\]

we have

\[
\frac{\partial c^*}{\partial k} = \frac{1}{\partial k/\partial c^*} = \frac{-1}{[1 - F(c^*)]^{N-1} + (N - 1)(\bar{b} - c^*)f(c^*)[1 - F(c^*)]^{N-2}} < 0
\]

Then it follows that

\[
\frac{db}{dk} = \frac{1}{[1 - F(c)]^{N-1}} + \frac{[1 - F(c^*)]^{N-1} \partial c^*}{[1 - F(c)]^{N-1} \partial k}
\]

\[
= \frac{1}{[1 - F(c)]^{N-1}} \left[ 1 - \frac{1 - F(c^*)}{1 - F(c^*) + (N - 1)(\bar{b} - c^*)f(c^*)} \right] > 0
\]

So a decrease in the bid-preparation cost will make the potential firms more likely to bid and the bidding firms bid more competitively.

In order to assess how the procurement auction model with the bid-preparation cost can fit the data we have observed, we use the structural estimates from Bayesian MCMC method to simulate the auctions in the data. For each auction in the data, we use the mean of last 15000 MCMC draws of \( \mu^* \) to estimate the mean of the project cost distribution. Then we draw an entry cost from the posterior density function of the entry cost, from which we can calculate the participation cut-off level, equilibrium...
probability of participation and equilibrium bidding function. Last we draw the project costs for each potential bidder in the auction and calculate the winning bid corresponding to the lowest project cost and the number of bidding firms. We simulate the winning bid and the number of bidding firms 500 times for each auction and then calculate the simulated mean of the winning bid as the procurement cost. Figure 14 displays the histograms of both observed/simulated procurement costs and the number of bidding firms. These histograms show a good model fitting ability.

In Section two we show that the bidding function is not necessarily a decreasing function of the number of potential bidders at an appropriate project cost. Two competing effects have been identified from an increase in the number of potential bidders: the competition effect, which always decreases the bid, and the entry cost effect. More specifically, the bidding function is

$$s(c|F,N,c^*) = c + \frac{\int_{c}^{c^*} [1 - F(t)]^{N-1} dt + (\bar{b} - c^*)[1 - F(c^*)]^{N-1}}{[1 - F(c)]^{N-1}}$$

from which we derive the the effect of the number of potential bidders on bid as follows

$$\frac{db}{dN} = \frac{\partial s(c|F,N,c^*)}{\partial N} + \frac{\partial s(c|F,N,c^*)}{\partial c^*} \frac{\partial c^*}{\partial N}$$

$$= \int_{c}^{c^*} \left[ \frac{1 - F(t)}{1 - F(c)} \right]^{N-1} \ln \left( \frac{1 - F(t)}{1 - F(c)} \right) dt$$

$$+ k \left[ \frac{1}{1 - F(c)} \right]^{N-1} \ln \left( \frac{1}{1 - F(c)} \right)$$

$$+ \frac{[1 - F(c^*)]^{N-1}}{[1 - F(c)]^{N-1}} \frac{[1 - F(c^*)]k \ln(1 - F(c^*))}{[1 - F(c^*)]^N + (N-1)kf(c^*)}$$

The first and third terms are negative, representing the competition effect, which results from the fact that more bidders with higher project cost advantage participate
in the auction. The second term is positive, representing the entry cost effect, which results from the fact that bidders need to recover the entry cost by submitting a higher bid. To have a more concrete analysis based on the discrete property of the variable $N$, we can separate these two competing effects as follows. Define $Total.Effect(N|c)$, $Comp.Effect(N|c)$ and $EntryCost.Effect(N|c)$ as the total, competition and entry cost effects from the increase in the number of potential bidder from $N-1$ to $N$ given the project cost $c$. Then

$$Total.Effect(N|c) = s(c|F, N, c_N^*) - s(c|F, N - 1, c_{N-1}^*)$$

$$= \frac{\int_c^{c_N}[1 - F(t)]^{N-1}dt + k}{[1 - F(c)]^{N-1}} - \frac{\int_c^{c_{N-1}}[1 - F(t)]^{N-2}dt + k}{[1 - F(c)]^{N-2}}$$

$$Comp.Effect(N|c) = \frac{\int_c^{c_N}[1 - F(t)]^{N-1}dt}{[1 - F(c)]^{N-1}} - \frac{\int_c^{c_{N-1}}[1 - F(t)]^{N-2}dt}{[1 - F(c)]^{N-2}}$$

and

$$EntryCost.Effect(N|c) = \frac{k}{[1 - F(c)]^{N-1}} - \frac{k}{[1 - F(c)]^{N-2}}$$

It is easy to show that competition effect is always negative, entry effect is always positive and

$$Total.Effect(N|c) = Comp.Effect(N|c) + EntryCost.Effect(N|c)$$

which could be both positive and negative, depending on the respective intensity of $Comp.Effect(N|c)$ and $EntryCost.Effect(N|c)$, which in turn depend on the number of potential bidders and the distribution of project cost. Given our parametric assumptions and Bayesian estimation results, in Figure 15 we show three effects for a representative procurement auction of median size in the data. The horizontal axis shows the the number of potential bidders. The vertical axis shows the level of three
effects defined above. We have used the simulated project cost mean and the bid-preparation cost mean to calculate the values of all effects. The competition effect curve is always below 0, which shows how much the bid will decrease from the partial competition effect of an unit increase in the number of potential bidders. The entry cost effect curve is always above 0, which shows how much the bid will increase from the partial entry cost effect of an unit increase in the number of potential bidders. The total effect curve is below 0 when the number of the potential bidders is small and is above 0 when the number of potential bidders is large. The figure shows when the number of potential bidders are large than 8, the increase in the number of potential bidders will have an adverse effect on the bid.

We can also quantify the effect of the number of potential bidders on the procurement cost\textsuperscript{30}, Although a bidding firm can possibly bid less aggressively when the number of potential firms increases, the expected procurement cost can still decrease with the number of potential bidders. This is because when the number of potential firms increases, the participation cut-off level will decrease. As a result, the bidding firms on average will have lower project costs. In Section two we have shown that the expected contract cost can be written as

\[
E[Cost|N] = NkF(c^*) + N \int_c^{c^*} \left[ 1 - F(t) \right]^{N-1} \left[ tf(t) + F(t) \right] dt + p_0 \left[ 1 - F(c^*) \right]^N
\]

Given our parametric assumptions and Bayesian estimation results, in Figure 16 we shows the effect of the number of potential bidders and the bid-preparation cost on the expected contract cost for the same representative procurement auction in the data. The figure shows when the number of potential bidders increases, the change

\textsuperscript{30}This analysis will depend on the cost for the procurer if the procurement project is not successfully contracted out. For simplicity, we assume the procurer’s project cost is the bid upper bound in our analysis.
in the expected procurement cost is always below 0, meaning that in this specific setting, more potential bidders lead to lower expected contract cost. The figure also shows that the expected procurement cost can decrease when the bid-preparation cost decreases. So if the government can effectively decrease the bid-preparation cost, it can benefit from a lower contract cost.

### 3.6 Conclusion

In this chapter we study the procurement auction models with entry cost. When there exists entry cost in the procurement auction, the potential bidders are confronted with both participation strategy and bidding strategy and the potential bidders are different from the actual bidders. It follows that the number of actual bidders are endogenously determined and depends on the project cost distribution. So the number of actual bidders can provide information for estimating the underlying project cost distribution. We establish the nonparametric identification for the model with the bid-preparation cost and the model with the information-gathering cost respectively. We also establish the condition to differentiate these two competing models. Due to the computational complexity, Bayesian method is used in the structural estimation of the procurement auction models with entry cost. We conduct an empirical analysis of the Michigan highway procurement market by using the procurement auction model with the bid-preparation cost.
APPENDIX A

PROOF OF PROPOSITIONS

Proof of Proposition 1. Auction theory has established the following result. Suppose in a $n$-firm procurement auction where $c$ follows the distribution $F_1(\cdot)$ on the support $[c, \tilde{c}]$, the equilibrium BNE bidding strategy is $s_1(c)$. Define $\tilde{c} = a + b \cdot c$ and let $F_2(\cdot)$ denote the distribution function of $\tilde{c}$ on the support $[a + b \cdot c, a + b \cdot \tilde{c}]$. Then in a $n$-firm procurement auction where $\tilde{c}$ follows the distribution function $F_2(\cdot)$, the equilibrium BNE bidding strategy can be written as $s_2(\tilde{c}) = a + b \cdot s_1(c)$ where $c = (\tilde{c} - a)/b$.

Given the result above, the bidding function in any auction with covariate vectors $X$ and $Z$ can be written as

$$b = s(c^*|X, Z, n, F) = X\beta + (Z\gamma) \cdot s_0(e|n, F)$$

where

$$s_0(e|n, F) = e + \frac{\int_{e}^{\tilde{e}} [1 - F(t)]^{n-1} dt}{[1 - F(e)]^{n-1}}$$

Define

$$g(n) = E[s_0(e|n, F)]$$

It follows that

$$E[b|X, n, \sigma] = X\beta + (Z\gamma) \cdot g(n)$$

84
Suppose \( g(n) = a + bn \). Given the fact that \( s_0(e|n, F) \) is a decreasing function of \( n \), \( g(n) \) must be a decreasing function of \( n \). It follows that \( b < 0 \). When \( n > 3 \),

\[
\begin{align*}
g(n) &= \int_\varepsilon^\pi \left[ c + \frac{\int_\varepsilon^\pi [1 - F(t)]^{n-1}dt}{1 - F(c)^{n-1}} \right] f(c)dc \\
&= \int_\varepsilon^\pi cf(c)dc + \int_\varepsilon^\pi \int_\varepsilon^\pi \frac{f(c)}{1 - F(c)^{n-1}} [1 - F(t)]^{n-1}dtdc \\
&= \int_\varepsilon^\pi \int_\varepsilon^t \frac{f(c)}{1 - F(c)^{n-1}} [1 - F(t)]^{n-1}dcdt \\
&= \int_\varepsilon^\pi [1 - F(t)]^{n-1} \left[ \frac{1}{n - 2} \frac{1}{1 - F(c)^{n-2}} \right] dt \\
&= \frac{1}{n - 2} \int_\varepsilon^\pi \left\{ [1 - F(t)] - [1 - F(t)]^{n-1} \right\} dt \\
&= -\frac{1}{n - 2} \int_\varepsilon^\pi t [1 - F(t)] - [1 - F(t)]^{n-1} \\
&= -\frac{1}{n - 2} \int_\varepsilon^\pi t [1 - F(t)]^{n-1} \\
&= -\frac{1}{n - 2} E[e_{n-1}^{(n-1)}]
\end{align*}
\]

where \( e_{n-1}^{(n-1)} \) is the smallest order statistic from \( n - 1 \) i.i.d. random samples. It follows that

\[
E[e_{n-1}^{(n-1)}] = -(n - 2)(a + bn) = -bn^2 + (2b - a)n + 2a
\]

So given that \( n \) is large enough, \( E[e_{(n-1:n-1)}] \) is increasing in \( n \), contradicting that fact that \( E[e_{n-1}^{(n-1)}] \) is a decreasing function of \( n \). A similar proof can be done for \( g(n) = a + b \cdot \ln(n) \). □

The following assumptions and notations are used in chapter 2.

(1) \( \{x_l, l = 1, \cdots, L\} \) are i.i.d. random variables from the distribution function \( H(\cdot) \) with density function \( h(\cdot) \).
(2) \( \{y_{il}, i = 1, \ldots, N\} \) are i.i.d. random variables from the conditional distribution function \( F(\cdot | x_l) \) with the corresponding conditional density function \( f(\cdot | x_l) \) for \( l = 1, \ldots, L \).

(3) The kernel function \( k(\cdot) \) is assumed to be a real, integrable, non-negative and even function with finite support and concentrated at the origin such that

\[
\int_R k(u)du = 1, \quad \int_R uk(u)du = 0, \quad \int_R u^2k(u)du < \infty
\]

and \( K(x) \) is defined as

\[
K(x) = \int_{-\infty}^{x} k(u)du
\]

(4) Regularity conditions are satisfied for the underlying distribution and density functions.

(5) Define

\[
\sigma_k^2 = \int w^2k(w)dw
\]
\[
R_0(k^2) = \int k^2(w)dw
\]
\[
R_2(k^2) = \int w^2k^2(w)dw
\]
\[
\varphi_1 = \int K(w)k(w)wdw
\]
\[
\varphi_3 = \int K(w)k(w)w^3dw
\]

**Lemma 12** Given assumptions (1)-(4) and \( Lh_f^2 \to \infty, h_f \to 0 \) as \( L \to \infty \), the kernel estimator of the joint density function defined by

\[
\hat{f}(x, y) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_f^2} \sum_{i=1}^{N} k \left( \frac{y - y_{il}}{h_f} \right) k \left( \frac{x - x_l}{h_f} \right)
\]
satisfies

\[
\sqrt{Lh_f^2} \left[ \hat{f}(x, y) - f(x, y) - \text{Asy.Bias}[\hat{f}(x, y)] \right] \overset{d}{\rightarrow} N \left( 0, \frac{R_0^2(k^2) f(y|x) h(x)}{N} \right)
\]

where

\[
\text{Asy.Bias}[\hat{f}(x, y)] = \frac{h_f^2 \sigma_k^2}{2} \left[ \frac{\partial^2 [f(y|x)h(x)]}{\partial x^2} + \frac{\partial^2 f(y|x)}{\partial y^2} h(x) \right] + O(h_f^4)
\]

**Proof.** Define

\[
q_{1l} = \frac{1}{Nh_f^3} \sum_{i=1}^{N} k \left( \frac{y - y_i}{h_f} \right) k \left( \frac{x - x_l}{h_f} \right)
\]

In the following calculation we will repeatedly use the lemmas in Hyndman, Bassertannyk and Grunwald (1996) and the assumption of conditional independence among \(\{y_i\}_{i=1}^{N}\). Note that

\[
E(q_{1l}|x_l) = E \left[ \frac{1}{Nh_f^3} \sum_{i=1}^{N} k \left( \frac{y - y_i}{h_f} \right) k \left( \frac{x - x_l}{h_f} \right) \right] = \frac{1}{h_f} k \left( \frac{x - x_l}{h_f} \right) E \left[ \frac{1}{Nh_f} \sum_{i=1}^{N} k \left( \frac{y - y_i}{h_f} \right) \right] = \frac{1}{h_f} k \left( \frac{x - x_l}{h_f} \right) E \left[ \frac{1}{h_f} k \left( \frac{y - y_i}{h_f} \right) \right] = \frac{1}{h_f} k \left( \frac{x - x_l}{h_f} \right) \left[ f(y|x_l) + \frac{h_f^2 \sigma_k^2}{2} \frac{\partial^2 f(y|x_l)}{\partial y^2} + O(h_f^4) \right]
\]

and

\[
\text{Var}(q_{1l}|x_l) = \frac{1}{h_f^4} k^2 \left( \frac{x - x_l}{h_f} \right) \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} k \left( \frac{y - y_i}{h_f} \right) \right] = \frac{1}{Nh_f^2} k^2 \left( \frac{x - x_l}{h_f} \right) \text{Var} \left[ \frac{1}{h_f} k \left( \frac{y - y_i}{h_f} \right) \right] = \frac{1}{Nh_f^2} k^2 \left( \frac{x - x_l}{h_f} \right) \times \left\{ f(y|x_l) \left[ \frac{R_0(k^2)}{h_f} - f(y|x_l) \right] + \frac{h_f R_2(k^2)}{2} \frac{\partial^2 f(y|x_l)}{\partial y^2} + O(h_f^2) \right\}
\]
It follows that

\[
E(q_{1l}) = \left[ f(y|x) + \frac{h_l^2 \sigma^2_k}{2} \frac{\partial^2 f(x)}{\partial y^2} \right] h(x) + \frac{h_l^2 \sigma^2_k}{2} \frac{\partial^2}{\partial x^2} [f(y|x)h(x)] + O(h_l^4)
\]

\[
= f(x, y) + \frac{\sigma^2_k}{2} \left\{ \frac{\partial^2}{\partial x^2} [f(y|x)h(x)] + \frac{\partial^2 f(y|x)}{\partial y^2} h(x) \right\} h_l^2 + O(h_l^4)
\]

\[
Var[E(q_{1l}|x_l)] = f^2(y|x)h(x) \left[ \frac{R_0(k^2)}{h_f} - h(x) \right] + O(h_f)
\]

and

\[
E[Var(q_{1l}|x_l)] = \frac{h(x)R_0(k^2)}{N h_f} \left\{ f(y|x) \left[ \frac{R_0(k^2)}{h_f} - f(y|x) \right] + \frac{h_f R_2(k^2)}{2} \frac{\partial^2 f(y|x)}{\partial y^2} \right\}
\]

\[+ \frac{R_0(k^2)R_2(k^2)}{2N} \frac{\partial [f(y|x)h(x)]}{\partial x} + O(h_f)
\]

Furthermore,

\[
Var(q_{1l}) = E[Var(q_{1l}|x_l)] + Var[E(q_{1l}|x_l)]
\]

\[= \frac{h(x)R_0(k^2)}{N h_f} \left\{ f(y|x) \left[ \frac{R_0(k^2)}{h_f} - f(y|x) \right] + \frac{h_f R_2(k^2)}{2} \frac{\partial^2 f(y|x)}{\partial y^2} \right\}
\]

\[+ \frac{R_0(k^2)R_2(k^2)}{2N} \frac{\partial [f(y|x)h(x)]}{\partial x} + f^2(y|x)h(x) \left[ \frac{R_0(k^2)}{h_f} - h(x) \right] + O(h_f)
\]

\[= -f^2(y|x)h^2(x) + \left[ 1 - \frac{1}{N} \right] R_0(k^2) h(x) f^2(y|x) \frac{1}{h_f} + R_0(k^2) f(y|x) h(x) \frac{1}{N h_f^2}
\]

\[+ \frac{R_0(k^2)R_2(k^2)}{2N} \frac{\partial^2 f(y|x)}{\partial y^2} h(x) + \frac{R_0(k^2)R_2(k^2)}{2N} \frac{\partial [f(y|x)h(x)]}{\partial x} + O(h_f)
\]

By using the Liapunov’s central limit theorem, we have that

\[
\frac{[\hat{f}(x, y) - E[\hat{f}(x, y)]]}{\sqrt{Var[\hat{f}(x, y)]}} \overset{d}{\to} N(0, 1)
\]

Since

\[
E[\hat{f}(x, y)] = E(q_{1l})
\]

\[
Var[\hat{f}(x, y)] = \frac{1}{L} Var(q_{1l})
\]

88
we have

\[
\frac{[\hat{f}(x, y) - E[\hat{f}(x, y)]]}{\sqrt{\text{Var}[\hat{f}(x, y)]}} = \frac{\sqrt{L}[\hat{f}(x, y) - f(x, y) - \text{Asy.Bias}[\hat{f}(x, y)]]}{\sqrt{\text{Var}(q_{11})}}
\]

\[
= \frac{\sqrt{Lh_f^2}[\hat{f}(x, y) - f(x, y) - \text{Asy.Bias}[\hat{f}(x, y)]]}{\sqrt{h_f^2\text{Var}(q_{11})}}
\]

\[
\sim \frac{\sqrt{Lh_f^2}[\hat{f}(x, y) - f(x, y) - \text{Asy.Bias}[\hat{f}(x, y)]]}{\sqrt{R_0^2(k^2)f(y|x)h(x)\frac{1}{N}}}
\]

under the assumptions that \(Lh_f^2 \to \infty, h_f \to 0\) as \(L \to \infty\). It follows that

\[
\frac{\sqrt{Lh_f^2}[\hat{f}(x, y) - f(x, y) - \text{Asy.Bias}[\hat{f}(x, y)]]}{\sqrt{R_0^2(k^2)f(y|x)h(x)\frac{1}{N}}} \xrightarrow{d} N(0, 1)
\]

i.e.,

\[
\sqrt{Lh_f^2} \left[ \hat{f}(x, y) - f(x, y) - \text{Asy.Bias}[\hat{f}(x, y)] \right] \xrightarrow{d} N \left( 0, \frac{R_0^2(k^2)f(y|x)h(x)}{N} \right)
\]

Lemma 13 Given assumptions (1)-(4) and \(Lh_F \to \infty, h_F \to 0\) as \(L \to \infty\), the kernel estimator of the conditional distribution function defined by

\[
\hat{F}(y|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_F} \sum_{i=1}^{N} 1(y_i \leq y) \left( \frac{x-x_i}{h_F} \right)
\]

satisfies

\[
E[\hat{F}(y|x)] = F(y|x) + \text{Asy.Bias}[\hat{F}(y|x)]
\]

and

\[
\text{Var}[\hat{F}(y|x)] = \frac{R_0(k^2)}{Lh_FNh(x)}F(y|x)[1 - F(y|x)]
\]

89
where

\[
\text{Asy.Bias}[\hat{F}(y|x)] = \frac{h^2_x \sigma^2_k}{2h(x)} \left[ \frac{\partial^2 F(y|x)}{\partial x^2} h(x) + 2 \frac{\partial F(y|x)}{\partial x} h'(x) \right] + O(h_F^4)
\]

**Proof.** Lemma 13 can be proved by using similar methods in Hyndman, Bashtannyk and Grunwald (1996).

**Lemma 14** Given assumptions (1)-(4) and \(Lh_y h_x \to \infty, h_y \to 0, h_x \to 0\) as \(L \to \infty\), the kernel estimator of the conditional density function defined by

\[
\hat{f}(y|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_y h_x} \sum_{i=1}^{N} k \left( \frac{y-y_i}{h_y} \right) k \left( \frac{x-x_l}{h_x} \right)
\]

satisfies

\[
\sqrt{Lh_y h_x} \left[ \hat{f}(y|x) - f(y|x) - \text{Asy.Bias}[\hat{f}(y|x)] \right] \overset{d}{\to} N \left( 0, \frac{R_0^2(k^2)f(y|x)}{Nh(x)} \right)
\]

where

\[
\text{Asy.Bias}[\hat{f}(y|x)] = \frac{h^2_y \sigma^2_k}{2} \frac{\partial^2 f(y|x)}{\partial y^2} + \frac{h^2_x \sigma^2_k}{2} \left[ \frac{h'(x)}{h(x)} \frac{\partial f(y|x)}{\partial x} + \frac{\partial^2 f(y|x)}{\partial x^2} \right]
\]

\[+ O(h_x^4) + O(h_y^4) + O(h_x^2 h_y^2)\]

**Proof.** Define

\[
\hat{f}(y, x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_y h_x} \sum_{i=1}^{N} k \left( \frac{y-y_i}{h_y} \right) k \left( \frac{x-x_l}{h_x} \right)
\]

By using similar method as that in Lemma 12, we can prove that

\[
\sqrt{Lh_y h_x} \left[ \hat{f}(y, x) - E[\hat{f}(y|x)] \right] \overset{d}{\to} N \left( 0, \frac{R_0^2(k^2)f(y|x)h(x)}{N} \right)
\]

where

\[
E[\hat{f}(y|x)] = f(y|x) + \text{Asy.Bias}[\hat{f}(y,x)]
\]
and

\[ \text{Asy.Bias}[\hat{f}(y, x)] = \frac{\sigma_k^2}{2} \frac{\partial^2 f(y|x)h(x)}{\partial x^2} h_x^2 + \frac{\sigma_k^2}{2} \frac{\partial^2 f(y|x)}{\partial y^2} h(x) h_y^2 + O(h_x^4) + O(h_y^4) + O(h_x^2 h_y^2) \]

Define

\[ \tilde{h}(x) = \frac{1}{L} \sum_{i=1}^{L} \frac{1}{h_x} k \left( \frac{x - x_i}{h_x} \right) \]

which is a \( \sqrt{Nh_x} \)-consistent estimator of \( h(x) \). Using Slutsky’s lemma, we have

\[ \sqrt{Nh_x} \left[ \frac{\hat{f}(y|x)}{\tilde{h}(x)} - \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} \right] \xrightarrow{d} N(0, \frac{R_0^2(k^2)f(y|x)}{Nh(x)}) \]

We can write

\[
\begin{align*}
\sqrt{Nh_x} \left[ \frac{\hat{f}(y|x)}{\tilde{h}(x)} - \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} \right] &= \sqrt{Nh_x} \left[ \frac{\hat{f}(y|x)}{\tilde{h}(x)} - \frac{E[\hat{f}(y|x)]}{\hat{h}(x)} \right] + \sqrt{Nh_x} \left[ \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} - \frac{E[\hat{f}(y|x)]}{\hat{h}(x)} \right] \\
&= \sqrt{Nh_x} \left[ \frac{\hat{f}(y|x)}{\tilde{h}(x)} - \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} \right] + \sqrt{Nh_x} \left[ E[\hat{f}(y|x)] - \hat{h}(x) \right] \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} \\
&= \sqrt{Nh_x} \left[ \frac{\hat{f}(y|x)}{\hat{h}(x)} \right] - \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} \right] + \sqrt{Nh_x} \left[ E[\hat{f}(y|x)] - \hat{h}(x) \right] \frac{E[\hat{f}(y|x)]}{\tilde{h}(x)} \\
\end{align*}
\]

The asymptotic distribution of the first term has been derived above and \( \sqrt{Nh_x}[E\tilde{h}(x) - \hat{h}(x)] \) converges to 0 consistently. So

\[ \sqrt{Nh_x} \left[ \frac{\hat{f}(y|x)}{\hat{h}(x)} - \frac{E[\hat{f}(y|x)]}{\hat{h}(x)} \right] \xrightarrow{d} N(0, \frac{R_0^2(k^2)f(y|x)}{Nh(x)}) \]

Using the fact that \( 1/(s + \delta) = 1/s - \delta/s^2 + o(\delta) \) when \( \delta \) approaches 0, we have

\[
\frac{E\hat{f}(y|x)}{E\hat{h}(x)} = \frac{f(y|x) + \text{Asy.Bias}[\hat{f}(y, x)]}{h(x) + \text{Asy.Bias}[\hat{h}(x)]} \\
= \frac{h_y^2 \sigma_k^2}{2} \frac{\partial^2 f(y|x)}{\partial y^2} + \frac{h_x^2 \sigma_k^2}{2} \left[ 2 h'(x) \frac{\partial f(y|x)}{\partial x} + \frac{\partial^2 f(y|x)}{\partial x^2} \right] \\
+ O(h_x^4) + O(h_y^4) + O(h_x^2 h_y^2) 
\]
Lemma 15  Given assumptions (1)-(4) and $L\to\infty$, $h_y\to 0$, $h_x\to 0$ as $L\to\infty$, the kernel estimator of the conditional survival function defined by

$$
\hat{S}(y|x) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{Nh_x} \sum_{i=1}^{N} K\left(\frac{y_i-y}{h_y}\right) k\left(\frac{x-x_l}{h_x}\right)
$$

satisfies

$$
E[\hat{S}(y|x)] = \hat{S}(y|x) + \text{Asy.Bias}[\hat{S}(y|x)]
$$

where

$$
\text{Asy.Bias}[\hat{S}(y|x)] = -\frac{h_y^2\sigma_k^2}{2} \frac{\partial f(y|x)}{\partial y} + \frac{h_x^2\sigma_k^2}{2h(x)} \left[ \frac{\partial^2 [S(y|x)]}{\partial x^2} h(x) + 2 \frac{\partial S(y|x)}{\partial x} h'(x) \right] + O(h_x^4) + O(h_y^4) + O(h_x^2h_y^2)
$$

Proof. We can prove lemma 15 by repeatedly using lemmas in Hyndman, Baskettannyk and Grunwald (1996) and the following results.

$$
E\left[ K\left(\frac{X-x}{a}\right) \right] = 1 - H(x) - \frac{1}{2} \sigma_k^2 h'(x)a^2 + O(a^4)
$$

$$
E\left[ K^2\left(\frac{X-x}{a}\right) \right] = 1 - H(x) - 2\varphi_1 h(x)a - \frac{1}{2} \sigma_k^2 h'(x)a^2 - \frac{1}{3} \varphi_3 h''(x)a^3 + O(a^4)
$$

$$
\text{Var}\left[ K\left(\frac{X-x}{a}\right) \right] = H(x) [1 - H(x)] - 2\varphi_1 h(x)a + \sigma_k^2 h'(x) \left[ \frac{1}{2} - H(x) \right] a^2 - \frac{1}{3} \varphi_3 h''(x)a^3 + O(a^4)
$$
Proof of Proposition 2. From Lemma 12, we have

\[
AMSE(\hat{f}(x, y)) = \left[ E\left[ \hat{f}(x, y) - f(x, y) \right] \right]^2 + \text{Var}[\hat{f}(x, y)]
\]

\[
= \frac{h_f^4 \sigma_k^4}{4} \left[ \frac{\partial^2 [f(y|x)h(x)]}{\partial x^2} + \frac{\partial^2 f(y|x)}{\partial y^2} h(x) \right]^2 + \frac{R_0(k^2)f(y|x)h(x)}{Nh_f^2} + O(h_f^6)
\]

It follows that

\[
IMSE(\hat{f}) = \int \int AMSE(\hat{f}(x, y))h(x)dydx = \frac{c_{11}}{Lh_f^2} + c_{12}h_f^4
\]

where

\[
c_{11} = \frac{R_0(k^2)}{N} \int \int f(y|x)h^2(x)dydx
\]

\[
c_{12} = \frac{\sigma_k^4}{4} \int \int \left[ \frac{\partial^2 [f(y|x)h(x)]}{\partial x^2} + \frac{\partial^2 f(y|x)}{\partial y^2} h(x) \right]^2 h(x)dydx
\]

From Lemma 13, we have

\[
AMSE(\hat{F}(y|x)) = \left[ E\left[ \hat{F}(|x) \right] - F(y|x) \right]^2 + \text{Var}[\hat{F}(y|x)]
\]

\[
= \frac{h_F^4 \sigma_k^4}{4h_F^2} \left[ \frac{\partial^2 [F(y|x)h(x)]}{\partial x^2} - F(y|x)F''(x) \right]^2
\]

\[
+ \frac{R_0(k^2)}{Nh_Fh(x)h_F} F(y|x)[1 - F(y|x)] + O(h_F^6)
\]

It follows that

\[
IMSE(\hat{F}) = \int \int AMSE(\hat{F}(y|x))h(x)dydx = \frac{c_{21}}{Lh_F} + c_{22}h_F^4
\]

where

\[
c_{21} = \frac{R_0(k^2)}{N} \int \int F(y|x)[1 - F(y|x)]dydx
\]

\[
c_{22} = \frac{\sigma_k^4}{4} \int \int \left[ \frac{\partial^2 [F(y|x)h(x)]}{\partial x^2} - F(y|x)F''(x) \right]^2 \frac{1}{h(x)} dydx
\]
Proof of Proposition 3. Since \( \hat{S}(y|x) \) is a \( \sqrt{Lh_x} \)-consistent estimator of \( S(y|x) \), using lemma 14 and Slutsky’s lemma, it follows

\[
\sqrt{Lh_x} h_x \left[ \frac{\hat{f}(y|x)}{\hat{S}(y|x)} - \frac{E[\hat{f}(y|x)]}{\hat{S}(y|x)} \right] \xrightarrow{d} N(0, \frac{R_0^2(k^2)f(y|x)}{Nh_x S^2(y|x)})
\]

We can write

\[
\sqrt{Lh_x} h_x \left[ \frac{\hat{f}(y|x)}{\hat{S}(y|x)} - \frac{E[\hat{f}(y|x)]}{\hat{S}(y|x)} \right] = \sqrt{Lh_x} h_x \left[ \frac{\hat{f}(y|x)}{\hat{S}(y|x)} - \frac{E[\hat{f}(y|x)]}{\hat{S}(y|x)} \right] + \sqrt{Lh_x} h_x \left[ \frac{E[\hat{f}(y|x)]}{\hat{S}(y|x)} - \frac{E[\hat{f}(y|x)]}{\hat{S}(y|x)} \right]
\]

The asymptotic distribution of the first term has been derived above and

\[
\sqrt{Lh_x} h_x [E[\hat{S}(y|x)] - \hat{S}(y|x)]
\]

can be shown to converge to 0 consistently. So

\[
\sqrt{Lh_x} h_x \left[ \frac{\hat{f}(y|x)}{\hat{S}(y|x)} - \frac{E[\hat{f}(y|x)]}{\hat{S}(y|x)} \right] \xrightarrow{d} N(0, \frac{R_0^2(k^2)f(y|x)}{Nh_x S^2(y|x)})
\]

By using Lemma 15 and the fact that \( 1/(s+\delta) = 1/s - \delta/s^2 + o(\delta) \) when \( \delta \) approaches 0, we have

\[
\frac{E[\hat{f}(y|x)]}{E[\hat{S}(y|x)]} = \frac{f(y|x) + \text{Asy.Bias}[\hat{f}(y|x)]}{S(y|x) + \text{Asy.Bias}[\hat{S}(y|x)]} = \lambda(y|x) + h_x^2 \sigma_k^2 \left[ \frac{\partial^2 f(y|x)h(x)}{S(y|x)h(x)} - \frac{f(y|x) \frac{\partial}{\partial x^2} [f(y|x)h(x)]}{S^2(y|x)h(x)} \right]
\]
\[
+ \frac{h_x^2 \sigma_k^2}{2} \left[ \frac{\partial^2 f(y|x)}{\partial y^2} \frac{1}{S(y|x)} + \frac{\partial f(y|x)}{\partial y} \frac{f(y|x)}{S^2(y|x)} \right]
\]
\[
+ O(h_x^4) + O(h_y^4) + O(h_x^2 h_y^2)
\]
It follows that

$$
\sqrt{L h_y h_x} \left[ \hat{\lambda}(y|x) - \lambda(y|x) - Asy.Bias[\hat{\lambda}(y|x)] \right] \overset{d}{\rightarrow} N(0, \frac{R_0^2(k^2)f(y|x)}{N h(x) S^2(y|x)})
$$

where

$$
Asy.Bias[\hat{\lambda}(y|x)] = \frac{h_x^2 \sigma_k^2}{2} \left[ \frac{\partial^2 [f(y|x)h(x)]}{S(y|x)h(x)} - \frac{f(y|x) \partial^2 [S(y|x)h(x)]}{S^2(y|x)h(x)} \right] + h_y^2 \sigma_k^4 \left[ \frac{\partial^2 f(y|x)}{\partial y^2} \frac{1}{S(y|x)} + \frac{\partial f(y|x)}{\partial y} \frac{f(y|x)}{S^2(y|x)} \right] + O(h_x^4) + O(h_y^4) + O(h_x^2 h_y^2)
$$

Then

$$
AMSE(x, y) = \left[ E \left[ \hat{\lambda}(y|x) \right] - \lambda(y|x) \right]^2 + Var[\hat{\lambda}(y|x)]
$$

$$
= C_1(x, y) \frac{1}{L h_x h_y} + C_2(x, y) h_x^4 + C_3(x, y) h_y^4 + C_4(x, y) h_x^2 h_y^2
$$

where

$$
C_1(x, y) = \frac{R_0^2(k^2)}{N} \frac{f(y|x)}{S^2(y|x)h(x)}
$$

$$
C_2(x, y) = \frac{\sigma_k^4}{4} \left[ \frac{\partial^2 [f(y|x)h(x)]}{S(y|x)h(x)} - \frac{\partial^2 [S(y|x)h(x)]f(y|x)}{S^2(y|x)h(x)} \right]^2
$$

$$
C_3(x, y) = \frac{\sigma_k^4}{4} \left[ \frac{\partial^2 f(y|x)}{\partial y^2} \frac{1}{S(y|x)} + \frac{\partial f(y|x)}{\partial y} \frac{f(y|x)}{S^2(y|x)} \right]^2
$$

$$
C_4(x, y) = \frac{\sigma_k^4}{2} \left[ \frac{\partial^2 [f(y|x)h(x)]}{S(y|x)h(x)} - \frac{\partial^2 [S(y|x)h(x)]f(y|x)}{S^2(y|x)h(x)} \right] \times \left[ \frac{\partial^2 f(y|x)}{\partial y^2} \frac{1}{S(y|x)} + \frac{\partial f(y|x)}{\partial y} \frac{f(y|x)}{S^2(y|x)} \right]
$$

Proof of Proposition 4. A marginal firm with the project cost equal to $c^*$ is indifferent between participating and not participating in the auction, i.e., its expected
profit $E\pi(c^*)$ will exactly cover the entry cost $k$. Note that a marginal firm can win if and only if it is the only bidder in the market, so

$$E\pi(c^*) = (\bar{b} - c^*)(1 - F(c^*))^{N-1}$$

Since $E\pi(c^*)$ is a strictly decreasing function of $c^*$ and $k \in (0, \bar{b} - \underline{c})$, there exists a unique solution $c^* \in (\underline{c}, \bar{b})$ such that $E\pi(c^*) = k$. For a bidder with project cost $c$ less than $c^*$, the expected profit conditional on the bidding strategy $s(\cdot)$ is

$$E\pi(b, c|s(\cdot)) = (b - c)[1 - F(s^{-1}(b))]^{N-1} - k$$

The first-order condition is characterized by the following differential equation

$$s'(c) = [s(c) - c](N - 1)\frac{f(c)}{1 - F(c)} \tag{A.1}$$

Combined with the boundary condition $s(c^*) = \bar{b}$, equation (A.1) will produce the unique equilibrium bidding function. To show this, note that when $k > 0$, $c^*$ is strictly less than $\bar{b}$. So it is easy to check that the Lipschitz condition\(^{31}\) for the differential equation above will always hold given our assumption that $f(c)$ is bounded away from 0. So the bidding strategy $s(\cdot, F, N, c^*)$ is uniquely determined on the interval $[\underline{c}, c^*]\(^{32}\).

More specifically, the bidding function can be written as

$$s(c, F, N, c^*) = c + \frac{\int_{\underline{c}}^{c} [1 - F(t)]^{N-1}dt + [1 - F(c^*)]^{N-1}(p_0 - c^*)}{[1 - F(c)]^{N-1}}$$

$$= c + \frac{\int_{\underline{c}}^{c} [1 - F(t)]^{N-1}dt + k}{[1 - F(c)]^{N-1}}$$

for $c \leq c^*$, where $c^*$ is determined by the equation $(\bar{b} - c^*)(1 - F(c^*))^{N-1} = k$ \(\blacksquare\)

\(^{31}\)Lipschitz condition is sufficient to guarantee the existence and uniqueness of the solution to a first-order differential equation with appropriate boundary conditions.

\(^{32}\)Note that when $k = 0$, $c^* = \bar{b}$, the model is reduced to the one without entry cost. The existence and uniqueness of the equilibrium bidding function have been established in the literature, though the Lipschitz condition fails in the case that $b = \bar{c}$.
Proof of Proposition 6. First we prove the necessity of C1, C2 and C3. A potential firm bids if and only if its project cost is less than or equal to the cut-off cost $c^*$, i.e. $c \leq c^*$ where $c^*$ is uniquely determined by the equation

$$(\bar{b} - c^*)[1 - F(c^*)]^{N-1} = k$$

From the assumption $k \in (0, \bar{b} - c)$, we know $c^* \in (\underline{c}, \bar{b})$, which means a potential firm will participate in the auction with probability $F(c^*) \in (0, 1)$. Given $N$ potential firms and the fact that the firms’ project costs are i.i.d., the number of bidding firms follows the binomial distribution with parameters $[N, F(c^*)]$. For any $(b_1, \ldots, b_n) \in R^+_n$, we have

$$\Pr(\tilde{b}_1 \leq b_1, \ldots, \tilde{b}_n \leq b_n) = \Pr(\tilde{c}_1 \leq s^{-1}(b_1), \ldots, \tilde{c}_n \leq s^{-1}(b_n))$$

$$= \prod_{i=1}^n \Pr(\tilde{c}_i \leq s^{-1}(b_i)) = \prod_{i=1}^n \frac{F(s^{-1}(b_i))}{F(c^*)}$$

i.e., conditional on $n$, $(\tilde{b}_1, \ldots, \tilde{b}_n)$ are i.i.d. as $F(s^{-1}(\cdot))/F(c^*)$.

Second we prove the sufficiency of C1, C2 and C3. Choose $\bar{c} \geq \bar{b}$ and let $\underline{c} = \xi(\bar{b}, G^*, N, F(c^*))$. Then we will prove that if C1, C2 and C3 hold, there exist a distribution $F(\cdot)$ on $[\underline{c}, \bar{c}]$ and $k \in (0, \bar{b} - \underline{c})$ that can rationalize $G^*(\cdot)$ and $\pi(\cdot)$ in a low-price sealed-bid auction with the bid-preparation cost $k$. Define a distribution function $F(\cdot)$ on $[\underline{c}, \bar{c}]$ such that $F(\cdot)$ can be written as $F(\cdot) = F(c^*)G^*(\xi^{-1}(\cdot, G^*, N, F(c^*)))$ on $[\underline{c}, \xi(\bar{b}, G^*, N, F(c^*))]$. The whole distribution $F(\cdot)$ on $[\underline{c}, \bar{c}]$ can be obtained by extension to $[\underline{c}, \bar{c}]$ in an absolute continuous fashion. Define $k = [\bar{b} - \xi(\bar{b}, G^*, N, F(c^*))][1 - F(c^*)]^{N-1}$. It is easy to check that $k \in (0, \bar{b} - \underline{c})$. Given the project cost distribution $F(\cdot)$ and the bid-preparation cost $k$ defined above, we can find the cut-off entry cost $\bar{c}^*$ and the equilibrium bidding strategy $s(\cdot, N, F(\cdot), \bar{c}^*)$. 
First, \( \tilde{c}^* \) must satisfy the equation
\[
(b - \tilde{c}^*)[1 - F(\tilde{c}^*)]^{N-1} = [\bar{b} - \xi(b, G^*, N, F(c^*))][1 - F(c^*)]^{N-1},
\]
and it is an obvious solution to the equation above. It is also the unique solution. Note that \( 1 - F(\tilde{c}^*) \) is nonincreasing in \( \tilde{c}^* \) since \( F(\tilde{c}^*) \) is a distribution function; \( 1 - F(\tilde{c}^*) > 0; \tilde{c} - \tilde{c}^* \) is strictly decreasing in \( \tilde{c}^* \). It follows that the solution must be unique because \( (b - \tilde{c}^*)[1 - F(\tilde{c}^*)]^{N-1} \) is strictly decreasing in \( \tilde{c}^* \). Also note that the entry probability for a potential firm is \( F(\tilde{c}^*) = F(c^*)G^*(\xi^{-1}(\tilde{c}^*, G^*, N, F(c^*)) = F(c^*) \). It follows \( \pi(\cdot) \) is the distribution of the number of bidding firms.

Second we need to show that the bid distribution is \( G^*(\cdot) \) given \( F(\cdot) \) and \( k \), i.e.,
\[
G^*(\cdot) = F(s^{-1}(\cdot, F, N, \tilde{c}^*)) / F(\tilde{c}^*)
\]
where \( s(\cdot, F, N, \tilde{c}^*) \) is the equilibrium strategy solving
\[
s'(c)[c](N - 1) f(c) = (N - 1)[c](N - 1) f(c)
\]
with the boundary condition \( s(\tilde{c}^*) = \bar{b} \). From the definition of \( F(\cdot) \), \( F(\cdot)/F(c^*) = G^*(\xi^{-1}(\cdot, G^*, N, F(c^*)) \) on \( [c, \xi(b, G^*, N, F(c^*))] \). Since \( F(c^*) = F(c^*) \), \( F(\cdot)/F(\tilde{c}^*) = G^*(\xi^{-1}(\cdot, G^*, N, F(c^*)) \). Then it suffices to show \( \xi^{-1}(\cdot, G^*, N, F(c^*)) \) is the equilibrium strategy, i.e., it solves
\[
\xi^{-1}(c, G^*, N, F(c^*)) = [\xi^{-1}(\cdot, G^*, N, F(c^*)) - c](N - 1) f(c)
\]
with the boundary condition \( \xi^{-1}(\tilde{c}^*, G^*, N, F(c^*)) = \bar{b} \). It is easy to check the boundary condition is satisfied. Since
\[
F(c^*) G^*(\xi^{-1}(\cdot, G^*, N, F(c^*)) \xi^{-1}(\cdot, G^*, N, F(c^*)) = \frac{f(c)}{1 - F(c)}
\]
it is equivalent to show that
\[
1 = [\xi^{-1}(c, G^*, N, F(c^*)) - c](N - 1) F(c^*) G^*(\xi^{-1}(c, G^*, N, F(c^*))
\]
By definition \( \xi(b, G^*, N, F(c^*)) = b - \frac{1 - G^*(b) F(c^*)}{(N - 1) g^*(b) F(c^*)} \). So the equation above is satisfied.

Last we prove the last part of the theorem. Since \( N \) and \( F(c^*) \) are identifiable from \( \pi(\cdot), \xi(\cdot, G^*, N, F(c^*)) \) is an unique function defined on \( [b, \bar{b}] \). From the proof
of necessity, we also know that \( \xi(\cdot, G^*, N, F(c^*)) = s^{-1}(\cdot, F, N, c^*) \) where \( F(c) = F(c^*) \) when \( c \) is evaluated at \( \bar{c} \) if \( F(\cdot) \) can rationalize both \( G^*(\cdot) \) and \( \pi(\cdot) \). Since the bid distribution \( G^*(\cdot) = \frac{F(s^{-1}(\cdot, F, N, \bar{c}))}{F(c^*)} \) it follows that \( F(\cdot) = F(c^*)G^*(s(\cdot, F, N, \bar{c})) = F(c^*)G^*(\xi^{-1}(\cdot, G^*, N, F(c^*))) \), which is uniquely defined on \( [\xi(h, G^*, N, F(c^*)), \xi(h, G^*, N, F(c^*))] \) \[ \square \]

**Proof of Proposition 7.** We know that a potential bidder will bid if and only if the ex ante expected payoff can recover the information-gathering cost. In equilibrium

\[
\pi(b, c|q) = \int_c^b [1 - qF(t)]^{N-1} dt
\]

So the ex ante expected payoff can be written as

\[
\int_{\xi}^b \pi(b, c|q)f(c)dc = \int_{\xi}^b \pi(b, c|q)f(c)dc = \int_{\xi}^b \int_c^b [1 - qF(t)]^{N-1} f(c)dt dc
\]

\[
= \int_{\xi}^b \int_c^t [1 - qF(t)]^{N-1} f(c)dc dt = \int_{\xi}^b \int_{\xi}^t [1 - qF(t)]^{N-1} F(t) dt
\]

Note that when \( k \geq \int_{\xi}^b F(t) dt \), the ex ante expected payoff can never recover the entry expense. Then no firm will participate, i.e., \( q^* = 0 \); when \( k \leq \int_{\xi}^b [1 - F(t)]^{N-1} F(t) dt \), the ex ante expected payoff can always recover the expense. Then every potential firm will participate, i.e., \( q^* = 1 \); when \( k \in [\int_{\xi}^b [1 - F(t)]^{N-1} F(t) dt, \int_{\xi}^b F(t) dt] \), \( q^* \) is determined by the unique solution to the equation \( \int_{\xi}^b [1 - qF(t)]^{N-1} F(t) dt = k \) \[ \square \]

**Proof of Proposition 9.** Suppose \([F(\cdot), k, N]\) is the underlying structural elements that generate the distribution \( G(\cdot) \) of the equilibrium bid and the distribution
Bin(\(N, P\)) of the number of bidding firms. By the equilibrium participation strategy, we know that

\[ b = c + \frac{[1 - qF(c)]s'(c)}{(N - 1)qf(c)} \]  

(A.2)

\[ P = q^*F(b) \]

and

\[ G(\cdot) = \frac{F(s^{-1}(b))}{F(b)} \]

Define an auction model structure without the information-gathering cost \([\hat{F}(\cdot), N]\) such that

\[ \hat{F}(\cdot) = q^*F(\cdot) \]

for \(c \in [c, \bar{c}]\). A potential firm will participate if and only if its project cost is less than \(\bar{b}\), then the number of bidding firms follows the distribution Bin(\(N, \hat{F}(\bar{b})\)) = Bin(\(N, P\)).

The BNE bidding function \(\hat{s}(\cdot)\) is characterized by the differential equation

\[ b = c + \frac{[1 - \hat{F}(c)]\hat{s}'(c)}{(N - 1)\hat{f}(c)} \]  

(A.3)

with the boundary condition \(\hat{s}(\bar{b}) = \bar{b}\). A comparison between equations (A.2) and (A.3) shows that \(\hat{s}(\cdot) = s(\cdot)\). Then it follows that

\[ \hat{G}(\cdot) = \frac{\hat{F}(s^{-1}(b))}{\hat{F}(\bar{b})} = \frac{q^*F(s^{-1}(b))}{q^*F(\bar{b})} = \frac{F(s^{-1}(b))}{F(\bar{b})} = G(\cdot) \]

\[ \Box \]

**Proof of Proposition 10.** First we prove the necessity of C1, C2 and C3. A potential firm bids with probability \(q^*\). Given \(N\) potential firms and the fact that the
firms’ private costs are i.i.d., the number of actual bidding firms follows the binomial distribution with parameter \((N, q^*)\). For any \((b_1, \cdots, b_n) \in \mathbb{R}_+^n\), we have

\[
\Pr(\tilde{b}_1 \leq b_1, \cdots, \tilde{b}_n \leq b_n) = \Pr(\tilde{c}_1 \leq s^{-1}(b_1), \cdots, \tilde{c}_n \leq s^{-1}(b_n)) = \prod_{i=1}^n \Pr(\tilde{c}_i \leq s^{-1}(b_i)) = \prod_{i=1}^n F(s^{-1}(b_i))
\]

i.e. conditional on \(n\), \((\tilde{b}_1, \cdots, \tilde{b}_n)\) are i.i.d. as \(F(s^{-1}(\cdot))\).

Second we prove the sufficiency of \(C1, C2\) and \(C3\). Choose \(\bar{c} = \tilde{b}\) and let \(\underline{c} = \xi(\bar{c}, G, N, q^*).\) Then we will prove that if \(C1, C2\) and \(C3\) hold, there exist a distribution \(F(\cdot)\) on \([\underline{c}, \bar{c}]\) and \(k \in (0, \bar{c} - \underline{c})\) that can rationalize \(G(\cdot)\) and \(\pi(\cdot)\) in a low-price sealed-bid auction with the information-gathering cost \(k\) and the reservation price \(p_0\).

Define a distribution function \(F(\cdot)\) on \([\underline{c}, \bar{c}]\) as \(F(\cdot) = G(\xi^{-1}(\cdot, G, N, q^*))\). Define \(k = \int_\underline{c}^\bar{c} [1 - q^*F(t)]^{N-1}F(t)dt\). It is easy to check that \(k \in (0, \bar{c} - \underline{c})\). Given the project cost distribution \(F(\cdot)\) and post-entry cost \(k\) defined above, we can find the entry probability \(\tilde{q}^*\) and the equilibrium bidding strategy \(\tilde{s}(\cdot, N, F(\cdot), \tilde{q}^*)\).

\[
\int_\underline{c}^\bar{c} [1 - q^*F(t)]^{N-1}F(t)dt = k = \int_\underline{c}^\bar{c} [1 - q^*F(t)]^{N-1}F(t)dt
\]

It follows that \(\tilde{q}^* = q^*\). So the \(\pi(\cdot)\) is the distribution of the number of bidding firms.

Second we need to show that the bid distribution is \(G^*(\cdot)\) given \(F(\cdot)\) and \(k\), i.e., \(G^*(\cdot) = F(\tilde{s}^{-1}(\cdot, N, F(\cdot), \tilde{q}^*))\) where \(\tilde{s}(\cdot, N, F(\cdot), \tilde{q}^*)\) solves

\[
s(c) = c + \frac{[1 - q^*F(c)]s'(c)}{(N - 1)q^*f(c)}
\]

with the boundary condition \(s(\bar{c}) = \bar{c}\). From the definition of \(F(\cdot)\),

\[
F(\cdot) = G(\xi^{-1}(\cdot, G, N, q^*))
\]

Then it suffices to show \(\xi^{-1}(\cdot, G^*, N, q^*)\) is the equilibrium strategy, i.e. it solves

\[
\xi^{-1}'(c, G^*, N, q^*) = c + \frac{[1 - q^*F(c)]s\xi^{-1}(c, G^*, N, q^*)}{(N - 1)q^*f(c)}
\]
with the boundary condition $\xi^{-1}(\bar{c}, G^*, N, F(c^*)) = \bar{c}$. It is easy to check the boundary condition is satisfied. By definition $\xi(b, G^*, N, F(c^*)) = b - \frac{1 - q^* G(b)}{(N - 1) q^* g(b)}$, the equation above is satisfied.

Last we prove the last part of the theorem. Since $N$ and $F(c^*)$ are identifiable from $\pi(\cdot)$, $\xi(\cdot, G^*, N, F(c^*))$ is an unique function defined on $[b, \bar{b}]$. From the proof of necessity, we also know that $\xi(\cdot, G^*, N, q^*) = s^{-1}(\cdot, F, N, q^*)$  

Proof of Proposition 11. In the model with the bid-preparation cost,

$$c = b - \frac{1 - F(c^*) G^{prep}(b)}{(N - 1) F(c^*) g^{prep}(b)}$$

Let $c$ approach $c^*$, given the boundary condition $\bar{b} = s(c^*)$, we have

$$c^* = \bar{b} - \frac{1 - F(c^*)}{(N - 1) F(c^*) g^{prep}(b)}$$

Since $c^* < \bar{b}$, it follows that $\lim_{b \uparrow \bar{b}} g^{prep}(b) < \infty$.

In the model with the information-gathering cost,

$$c = b - \frac{1 - q^* G^{info}(b)}{(N - 1) q^* g^{info}(b)}$$

Let $c$ approach $\bar{c}$, given the boundary condition $\bar{b} = \bar{c} = s(\bar{c})$, we have

$$\bar{c} = \bar{c} - \frac{1 - q^*}{(N - 1) q^* g^{prep}(b)}$$

It follows that $\lim_{b \uparrow \bar{b}} g^{prep}(b) = +\infty$  

102
APPENDIX B

DETAILS OF METROPOLIS-WITHIN-GIBBS ALGORITHM

The Metropolis-within-Gibbs algorithm consists of the following four steps:

(1) Draw of $(\mu_l, q_l^*)$ for $l = 1, \cdots, L$. $q_l^{new}$ is drawn from the the uniform distribution on $(0, 1)$ and $\mu_l^{new}$ is drawn from the normal proposal density as $N(\mu_l^{old}, \nu)$. After $(\mu_l, q_l^*)$ is drawn, we evaluate the unnormalized joint density of all bids $\{b\}_{l}^{n_l}$ and the number of bidders $n_l$ in auction $l$, which involves the computation of the inverse bidding function at all bids. Since the proposal density is symmetric between $(\mu_l^{old}, q_l^{old})$ and $(\mu_l^{new}, q_l^{new})$, $(\mu_l^{new}, q_l^{new})$ is accepted with the probability

$$\min \left[ \frac{\pi(\mu_l^{new}, q_l^{new}|\beta, \theta, d_l, \sigma_l^2, b_l, n_l)}{\pi(\mu_l^{old}, q_l^{old}|\beta, \theta, d_l, \sigma_l^2, b_l, n_l)}, 1 \right].$$

(2) Draw of $\beta$. Given the results in step (1), $\beta^{new}$ is drawn from the distribution

$$N \left( \left[ D_\beta + \sum_{l=1}^{L} \frac{x_l'x_l}{\sigma_l^2} \right]^{-1} [D_\beta \beta_0 + \sum_{l=1}^{L} \frac{x_l'(x_l - d_l)}{\sigma_l^2}], [D_\beta + \sum_{l=1}^{L} \frac{x_l'x_l}{\sigma_l^2}]^{-1} \right).$$

(3) Draw of $\theta$. Given the results in step (1) and (2), we use a similar method adopted by Li and Zheng (2006) and originated from Chib et al (1998). Let $\hat{\theta}$ denote the mode of the full conditional density. The proposal density is selected to be multivariate normal distribution $N(\theta^{new}|\hat{\theta} - (\theta^{old} - \hat{\theta}), \tau V)$. The covariance matrix
$V$ is set to be $V = -H^{-1}$ where $H$ is the value of Hessian matrix $\ln[f(\theta|\mu, q^*)]$ at $\hat{\theta}$. 

$\theta^{new}$ is accepted with the probability

$$\min \left[ \frac{f(\theta^{new}|\mu, q^*)}{f(\theta^{old}|\mu, q^*)}, 1 \right].$$

(4) Update of $(d_l, \sigma_l^2)$ for $l = 1, \cdots, L$. Given the results from previous steps, we can obtain the data of the unnormalized error terms characterizing the unobserved heterogeneity. Then we can use the data to update the number of clusters, the mean and variance of each cluster and the cluster identity from which $(d_l, \sigma_l^2)$ is drawn from for $l = 1, \cdots, L$. In this step we use the Semiparametric Bayesian density estimation method. More specifically, the unnormalized error terms come from a Dirichlet mixture of normals, i.e.,

$$e_l|d_l, \sigma_l^2 \sim N(d_l, \sigma_l^2) \text{ for } l = 1, \cdots, L$$

$$(d_l, \sigma_l^2)|\alpha, G_0 \sim DP(\alpha G_0)$$

where $\alpha$ is the precision parameter and $G_0$ is the baseline distribution. In the model specification, $G_0$ is the conjugate normal-inverted-Wishart distribution

$$G_0 = N(\mu|m_1, (1/k_0)\Sigma)IW(\Sigma|v_1, \psi_1)$$

and independent hyperpriors are assumed as follows

$$\alpha|a_0, b_0 \sim Gamma(a_0, b_0)$$

$$m_1|m_2, s_2 \sim N(m_2, s_2)$$

$$k_0|\tau_1, \tau_2 \sim Gamma(\tau_1/2, \tau_2/2)$$
\[ \psi_1 | v_2, \psi_1 \sim IW(v_2, \psi_1) \]

For details of this algorithm, please refer to Escobar and West (1995) and Li and Zheng (2005).

The programs from step (1) to step (3) are written in MATLAB 7.1. The update in step (4) is implemented by the DPpackage 1.0-2\(^{33}\). Since DPpackage is written in R, MATLAB R-link\(^{34}\) is used to call R functions from within MATLAB. On the platform of Pentium 4 2GHz CPU and 1G RAM, it took 36 hours to finish 20000 random draws for latent variables and parameters. Calculation of the inverse bidding function, which is a root-finding problem, is the most time-consuming part of the estimation.

\(^{33}\)DPpackage 1.0-2, written by Alejandro Jara Vallejos, allows the user to perform Bayesian inference via simulation from the posterior distributions for models considering Dirichlet Process and Polya Tree priors.

\(^{34}\)MATLAB R-link, written by Robert Henson, is a COM based interface that allows users to call the statistical package R from within MATLAB.
### APPENDIX C

#### TABLES

<table>
<thead>
<tr>
<th></th>
<th>( h_g )</th>
<th>( h_G )</th>
<th>( h_x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 0 (x=10)</td>
<td>1.9470 (0.1114)</td>
<td>1.5466 (0.0885)</td>
<td>2.5871 (0.1507)</td>
</tr>
<tr>
<td>Method 1 (x=10)</td>
<td>0.5529 (0.0096)</td>
<td>0.7717 (0.0198)</td>
<td>0.7717 (0.0198)</td>
</tr>
<tr>
<td>Method 2 (x=10)</td>
<td>0.7640 (0.0071)</td>
<td>( 0 )</td>
<td>0.3104 (0.0136)</td>
</tr>
</tbody>
</table>

Table 1: Bandwidth Used in Estimating Inverse Bidding Functions*

*: Each entry gives the mean and std var of 50 bandwidth used in simulation

<table>
<thead>
<tr>
<th></th>
<th>Method 0</th>
<th>Method 1</th>
<th>Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMSE (x=8)</td>
<td>3.0500e-2 (1.0500e-2)</td>
<td>3.4000e-3 (1.5000e-3)</td>
<td>2.8000e-3 (3.3000e-3)</td>
</tr>
<tr>
<td>CIMSE (x=8)</td>
<td>1.1000e-3 (7.5303e-4)</td>
<td>5.8023e-5 (5.3821e-5)</td>
<td>3.1719e-5 (3.9437e-5)</td>
</tr>
<tr>
<td>CIMSE (x=10)</td>
<td>8.0266e-4 (6.0030e-4)</td>
<td>5.1157e-5 (4.7809e-5)</td>
<td>5.0880e-5 (7.4185e-5)</td>
</tr>
<tr>
<td>CIMSE (x=12)</td>
<td>6.9195e-4 (5.6404e-4)</td>
<td>4.9466e-5 (4.2006e-5)</td>
<td>4.0546e-5 (5.7781e-5)</td>
</tr>
</tbody>
</table>

*: IMSE and CIMSE are calculated under \( M1 = M2 = 30 \)

Table 2: CMISE/MSE of Estimated Inverse Bidding Functions*
Table 3: Explanation of Variables in Ohio Highway Procurement Market

<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LnNBid</td>
<td>log of the bid divided by 100000</td>
</tr>
<tr>
<td>LnNEE</td>
<td>log of the engineer’s estimate divided by 100000</td>
</tr>
<tr>
<td>LnDays</td>
<td>log of the estimated working days</td>
</tr>
<tr>
<td>ProjType(D)</td>
<td>1 for 2-lane resurfacing and 0 for 4-lane resurfacing</td>
</tr>
<tr>
<td>LnDis</td>
<td>log of 1+the distance from county of the firm to county of the project</td>
</tr>
<tr>
<td>WT(D)</td>
<td>1 if the firm has unfinished project in the county where the new project is located and 0 otherwise</td>
</tr>
<tr>
<td>LnRMDis</td>
<td>log of 1+the minimum distance of other participating rival firms</td>
</tr>
<tr>
<td>LnNEEPday</td>
<td>log of the e.e. divided by 100000×working days</td>
</tr>
<tr>
<td>Entry</td>
<td>1 if a potential firm submits a bid and 0 otherwise</td>
</tr>
<tr>
<td>FirmType(D)</td>
<td>1 if the firm is of fring type and 0 if the firm is of regular type</td>
</tr>
</tbody>
</table>

Table 4: Summary Statistics of Main Variables (monetary unit:$100000)

<table>
<thead>
<tr>
<th>Variable</th>
<th># of obs</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bids</td>
<td>3771</td>
<td>18.94</td>
<td>39.83</td>
<td>0.0735</td>
<td>729.81</td>
</tr>
<tr>
<td>Wining bids</td>
<td>1259</td>
<td>15.36</td>
<td>34.66</td>
<td>0.0735</td>
<td>691.49</td>
</tr>
<tr>
<td>E.E.</td>
<td>1259</td>
<td>16.40</td>
<td>36.58</td>
<td>0.16</td>
<td>735.70</td>
</tr>
<tr>
<td># of bidders</td>
<td>1259</td>
<td>3.00</td>
<td>1.56</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td># of R bidders</td>
<td>1259</td>
<td>1.68</td>
<td>1.04</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td># of F bidders</td>
<td>1259</td>
<td>1.32</td>
<td>1.68</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 5: Division of Firms by the Number of Total Bids

<table>
<thead>
<tr>
<th># of total bids per firm</th>
<th># of firms</th>
<th># of all bids</th>
<th># of winning bids</th>
<th>Freq. of bidding</th>
<th>Freq. of winning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10</td>
<td>105</td>
<td>269</td>
<td>52</td>
<td>0.0713</td>
<td>0.0413</td>
</tr>
<tr>
<td>11-126</td>
<td>44</td>
<td>1495</td>
<td>427</td>
<td>0.3964</td>
<td>0.3392</td>
</tr>
<tr>
<td>127-472</td>
<td>8</td>
<td>2007</td>
<td>780</td>
<td>0.5322</td>
<td>0.6195</td>
</tr>
<tr>
<td>Total</td>
<td>157</td>
<td>3771</td>
<td>1259</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Project Size</td>
<td># of Proj</td>
<td>Mean # of F</td>
<td>Mean # of R</td>
<td>Mean Prob of Winning for F</td>
<td>Mean Freq of Winning for F</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>(0,0.5m]</td>
<td>177</td>
<td>2.0452</td>
<td>1.6497</td>
<td>0.5164</td>
<td>0.5593</td>
</tr>
<tr>
<td>(0.5m,1m]</td>
<td>170</td>
<td>1.7941</td>
<td>1.7529</td>
<td>0.4816</td>
<td>0.4765</td>
</tr>
<tr>
<td>(1m,73.57m]</td>
<td>219</td>
<td>2.1963</td>
<td>1.8265</td>
<td>0.5055</td>
<td>0.4155</td>
</tr>
</tbody>
</table>

F: fringe firm; R: regular firm

Table 6: Division of Projects by Size

Table 8: Estimated Markups in Different Subgroups

<table>
<thead>
<tr>
<th>Project Category</th>
<th>(\frac{\text{COST}}{\text{EE}}) of Fringe Firms</th>
<th>(\frac{\text{COST}}{\text{EE}}) of Regular Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Projects</td>
<td>0.9171 (0.1770)</td>
<td>0.8976 (0.1586)</td>
</tr>
<tr>
<td>Small Projects</td>
<td>0.8860 (0.1694)</td>
<td>0.9059 (0.1737)</td>
</tr>
<tr>
<td>((EE \leq 4 \cdot 10^5))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium Projects</td>
<td>0.9163 (0.1807)</td>
<td>0.8870 (0.1432)</td>
</tr>
<tr>
<td>(4 \cdot 10^5 &lt; EE \leq 10^6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large Projects</td>
<td>0.9366 (0.1756)</td>
<td>0.8813 (0.1565)</td>
</tr>
<tr>
<td>(EE &gt; 10^6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: Each entry gives the mean and standard of the estimates

Table 9: Summary Statistics of the Normalized Cost
<table>
<thead>
<tr>
<th>Group</th>
<th>All Firms</th>
<th>Regular Firms</th>
<th>Fringe Firms</th>
<th>All Firms</th>
<th>Regular Firms</th>
<th>Fringe Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>Probit</td>
<td>Probit</td>
<td>Probit</td>
</tr>
<tr>
<td>Dep. Var.</td>
<td>LnNBid</td>
<td>LnNBid</td>
<td>LnNBid</td>
<td>Entry</td>
<td>Entry</td>
<td>Entry</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.9877</td>
<td>0.9872</td>
<td>0.9884</td>
<td>0.4055</td>
<td>0.2927</td>
<td>0.2738</td>
</tr>
<tr>
<td>Number of Var.</td>
<td>3582</td>
<td>1950</td>
<td>1632</td>
<td>167990</td>
<td>9630</td>
<td>158360</td>
</tr>
<tr>
<td>Exp. Var.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.2360**</td>
<td>-0.2306</td>
<td>0.0463</td>
<td>-0.3309**</td>
<td>2.1916**</td>
<td>-2.4232**</td>
</tr>
<tr>
<td></td>
<td>(0.0280)</td>
<td>(0.0341)</td>
<td>(0.1461)</td>
<td>(0.1061)</td>
<td>(0.2105)</td>
<td>(0.1292)</td>
</tr>
<tr>
<td>LnNEE</td>
<td>0.9956**</td>
<td>0.9893**</td>
<td>1.0061**</td>
<td>0.0416**</td>
<td>0.0352</td>
<td>0.0652**</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td>(0.0038)</td>
<td>(0.0051)</td>
<td>(0.0180)</td>
<td>(0.0329)</td>
<td>(0.0219)</td>
</tr>
<tr>
<td>LnDays</td>
<td>0.0247**</td>
<td>0.0291**</td>
<td>0.0166*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0052)</td>
<td>(0.0064)</td>
<td>(0.0089)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProjType(D)</td>
<td>-0.0050</td>
<td>-0.0045</td>
<td>-0.0093</td>
<td>0.1112**</td>
<td>-0.0044</td>
<td>0.1581**</td>
</tr>
<tr>
<td></td>
<td>(0.0060)</td>
<td>(0.0075)</td>
<td>(0.0101)</td>
<td>(0.0255)</td>
<td>(0.0463)</td>
<td>(0.0314)</td>
</tr>
<tr>
<td>LnDis</td>
<td>0.0129**</td>
<td>0.0153**</td>
<td>0.0107**</td>
<td>-0.4122**</td>
<td>-0.9241**</td>
<td>-0.3659**</td>
</tr>
<tr>
<td></td>
<td>(0.0019)</td>
<td>(0.0028)</td>
<td>(0.0027)</td>
<td>(0.0067)</td>
<td>(0.0267)</td>
<td>(0.0072)</td>
</tr>
<tr>
<td>WT(D)</td>
<td>-0.0045</td>
<td>-0.0051</td>
<td>-0.0017</td>
<td>1.5250**</td>
<td>1.4283**</td>
<td>1.5165**</td>
</tr>
<tr>
<td></td>
<td>(0.0055)</td>
<td>(0.0065)</td>
<td>(0.0101)</td>
<td>(0.0396)</td>
<td>(0.0610)</td>
<td>(0.0536)</td>
</tr>
<tr>
<td># of Bidders</td>
<td>-0.0036**</td>
<td>-0.0007</td>
<td>-0.0057**</td>
<td>0.1622**</td>
<td>0.1440**</td>
<td>0.1795**</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0024)</td>
<td>(0.0026)</td>
<td>(0.0061)</td>
<td>(0.0129)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>LnRMDis</td>
<td>0.0029*</td>
<td>0.0024</td>
<td>0.0034</td>
<td>0.1172**</td>
<td>0.1215**</td>
<td>0.1045**</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0019)</td>
<td>(0.0025)</td>
<td>(0.0063)</td>
<td>(0.0112)</td>
<td>(0.0077)</td>
</tr>
<tr>
<td>FirmType(D)</td>
<td>0.1439</td>
<td></td>
<td></td>
<td>-1.6602**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1294)</td>
<td></td>
<td></td>
<td>(0.0212)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LnNEEPDay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* : 10% significant  ** : 5% significant

Table 7: OLS and Probit Regression Results
### Table 10: OLS Regression for Recovered Cost (Dep Var: LnCost)

<table>
<thead>
<tr>
<th>Exp. Var.</th>
<th>Estimate</th>
<th>Std</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.2996*</td>
<td>0.0632</td>
<td>-4.74</td>
</tr>
<tr>
<td>LnEE</td>
<td>1.0293**</td>
<td>0.0096</td>
<td>108.79</td>
</tr>
<tr>
<td>D</td>
<td>0.1711**</td>
<td>0.0735</td>
<td>2.33</td>
</tr>
<tr>
<td>D*LnEE</td>
<td>-0.0287**</td>
<td>0.0109</td>
<td>-2.62</td>
</tr>
</tbody>
</table>

R-Squared 0.9630

D: =1 for regular firms; =0 for fringe firms

*: 10% significant, **: 5% significant

### Table 11: Summary Statistics for Michigan Highway Procurement Market

<table>
<thead>
<tr>
<th>Variables</th>
<th>Explanation</th>
<th>Obs</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>StD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EE</td>
<td>Engineer’s Estimate</td>
<td>553</td>
<td>102280</td>
<td>6910600</td>
<td>1150100</td>
<td>1281100</td>
</tr>
<tr>
<td>Bid</td>
<td>Observed bid</td>
<td>2252</td>
<td>79190</td>
<td>11739000</td>
<td>1284000</td>
<td>1466200</td>
</tr>
<tr>
<td>Day</td>
<td>Planned working days</td>
<td>553</td>
<td>5</td>
<td>1135</td>
<td>122</td>
<td>134</td>
</tr>
<tr>
<td>Mile</td>
<td>Mileage of the project</td>
<td>553</td>
<td>0.04</td>
<td>47.1</td>
<td>3.4091</td>
<td>4.3770</td>
</tr>
<tr>
<td>Potential</td>
<td>Number of plan holders</td>
<td>553</td>
<td>1</td>
<td>22</td>
<td>6.5</td>
<td>4.4</td>
</tr>
<tr>
<td>Actual</td>
<td>Number of actual bidders</td>
<td>553</td>
<td>1</td>
<td>13</td>
<td>4</td>
<td>2.4</td>
</tr>
<tr>
<td>Nbid</td>
<td>Normalized bid (Bid/EE)</td>
<td>2252</td>
<td>0.3458</td>
<td>1.9355</td>
<td>1.0617</td>
<td>0.1630</td>
</tr>
<tr>
<td>Entry</td>
<td>Actual/Potential</td>
<td>553</td>
<td>0.1667</td>
<td>1</td>
<td>0.7311</td>
<td>0.2199</td>
</tr>
<tr>
<td>LogEE</td>
<td>Log(EE)</td>
<td>553</td>
<td>11.5355</td>
<td>15.7486</td>
<td>13.5012</td>
<td>0.9371</td>
</tr>
<tr>
<td>LogBid</td>
<td>Log(Bid)</td>
<td>2252</td>
<td>11.2796</td>
<td>16.2785</td>
<td>13.6066</td>
<td>0.9404</td>
</tr>
</tbody>
</table>

Table 11: Summary Statistics for Michigan Highway Procurement Market
### Table 12: Regression Analysis on Number of Actual Bidders

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. Var.</td>
<td>LogBid</td>
<td>LogBid (D)</td>
<td>LogBid</td>
<td>LogBid (D)</td>
<td>NBid</td>
<td>NBid</td>
<td></td>
</tr>
<tr>
<td>LogEE</td>
<td>1.0157*</td>
<td>1.0383*</td>
<td>1.0150*</td>
<td>1.0383*</td>
<td>0.0457</td>
<td>0.0457</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0046)</td>
<td>(0.0492)</td>
<td>(0.0046)</td>
<td>(0.0492)</td>
<td>(0.0551)</td>
<td>(0.0551)</td>
<td></td>
</tr>
<tr>
<td>LogDay</td>
<td>-0.0085**</td>
<td>-0.0027</td>
<td>-0.0091*</td>
<td>-0.0027</td>
<td>-0.0013</td>
<td>-0.0013</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0044)</td>
<td>(0.0315)</td>
<td>(0.0045)</td>
<td>(0.0315)</td>
<td>(0.0353)</td>
<td>(0.0353)</td>
<td></td>
</tr>
<tr>
<td>LogMile</td>
<td>-0.0101*</td>
<td>-0.0431</td>
<td>-0.0089*</td>
<td>-0.0431</td>
<td>-0.0533</td>
<td>-0.0533</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0037)</td>
<td>(0.0302)</td>
<td>(0.0038)</td>
<td>(0.0302)</td>
<td>(0.0339)</td>
<td>(0.0339)</td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>-0.0001</td>
<td>0.0183</td>
<td>0.0059</td>
<td>0.0251</td>
<td>0.0250</td>
<td>0.0338</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0903)</td>
<td>(0.0055)</td>
<td>(0.1131)</td>
<td>(0.1012)</td>
<td>(0.1267)</td>
<td></td>
</tr>
<tr>
<td>Actual^2</td>
<td>-0.0004</td>
<td>-0.0023</td>
<td>-0.0029</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0076)</td>
<td>(0.0085)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.1225*</td>
<td>-0.3471</td>
<td>-0.1265*</td>
<td>-0.3516</td>
<td>0.5628</td>
<td>0.5569</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0530)</td>
<td>(0.6627)</td>
<td>(0.0531)</td>
<td>(0.6639)</td>
<td>(0.7424)</td>
<td>(0.7437)</td>
<td></td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.9739</td>
<td>0.9879</td>
<td>0.9739</td>
<td>0.9879</td>
<td>0.4930</td>
<td>0.4930</td>
<td></td>
</tr>
</tbody>
</table>

### Table 13: Regression Analysis on Number of Potential Bidders

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. Var.</td>
<td>LogBid</td>
<td>LogBid (D)</td>
<td>LogBid</td>
<td>LogBid (D)</td>
<td>NBid</td>
<td>NBid</td>
</tr>
<tr>
<td>LogEE</td>
<td>1.0132*</td>
<td>0.9958*</td>
<td>1.0127*</td>
<td>1.0080*</td>
<td>-0.0051</td>
<td>0.0118</td>
</tr>
<tr>
<td></td>
<td>(0.0047)</td>
<td>(0.0341)</td>
<td>(0.0048)</td>
<td>(0.0365)</td>
<td>(0.0382)</td>
<td>(0.0409)</td>
</tr>
<tr>
<td>LogDay</td>
<td>-0.0095*</td>
<td>-0.0201</td>
<td>-0.0100*</td>
<td>0.0398</td>
<td>-0.0232</td>
<td>0.0486</td>
</tr>
<tr>
<td></td>
<td>(0.0044)</td>
<td>(0.0309)</td>
<td>(0.0045)</td>
<td>(0.0393)</td>
<td>(0.0346)</td>
<td>(0.0441)</td>
</tr>
<tr>
<td>LogMile</td>
<td>-0.0065**</td>
<td>-0.0243</td>
<td>-0.0055</td>
<td>-0.0348</td>
<td>-0.0310</td>
<td>-0.0448</td>
</tr>
<tr>
<td></td>
<td>(0.0038)</td>
<td>(0.0257)</td>
<td>(0.0041)</td>
<td>(0.0305)</td>
<td>(0.0288)</td>
<td>(0.0341)</td>
</tr>
<tr>
<td>Potential</td>
<td>0.0013**</td>
<td>-0.0111*</td>
<td>0.0032</td>
<td>-0.0497*</td>
<td>-0.0122*</td>
<td>-0.0579*</td>
</tr>
<tr>
<td></td>
<td>(0.0008)</td>
<td>(0.0048)</td>
<td>(0.0029)</td>
<td>(0.0208)</td>
<td>(0.0053)</td>
<td>(0.0233)</td>
</tr>
<tr>
<td>Potential^2</td>
<td>-0.0001</td>
<td>0.0017*</td>
<td></td>
<td>0.0020*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0008)</td>
<td></td>
<td>(0.0009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0966**</td>
<td>0.3384</td>
<td>-0.0954**</td>
<td>0.0410</td>
<td>1.3900*</td>
<td>1.0014*</td>
</tr>
<tr>
<td></td>
<td>(0.0546)</td>
<td>(0.4624)</td>
<td>(0.0546)</td>
<td>(0.4098)</td>
<td>(0.5180)</td>
<td>(0.4590)</td>
</tr>
<tr>
<td>Adjusted R^2</td>
<td>0.9739</td>
<td>0.9879</td>
<td>0.9739</td>
<td>0.9879</td>
<td>0.4930</td>
<td>0.4930*</td>
</tr>
</tbody>
</table>
### Table 14: Summary of the Effects of N/n on LogBid and Nbid

<table>
<thead>
<tr>
<th>Effect</th>
<th>LogBid</th>
<th>Nbid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>Actual-d</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>Actual</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>Actual^2</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>Actual-d</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>Actual^2-d</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>Potential</td>
<td>+</td>
<td>Yes</td>
</tr>
<tr>
<td>Potential-d</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>Potential</td>
<td>+</td>
<td>No</td>
</tr>
<tr>
<td>Potential^2</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>Potential-d</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>Potential^2-d</td>
<td>+</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Table 15: Regression for the Dependent Variable: Actual Number of Bidders

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Var</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogEE</td>
<td>-0.0214</td>
<td>0.0313</td>
<td>-0.68</td>
</tr>
<tr>
<td>LogDay</td>
<td>-0.0183</td>
<td>0.0262</td>
<td>-0.70</td>
</tr>
<tr>
<td>LogMile</td>
<td>-0.0158</td>
<td>0.0292</td>
<td>-0.54</td>
</tr>
<tr>
<td>LogPotential</td>
<td>0.7046*</td>
<td>0.0421</td>
<td>16.74</td>
</tr>
<tr>
<td>Constant</td>
<td>0.5167</td>
<td>0.3520</td>
<td>1.47</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-951.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 16: Regression for the Dependent Variable: Entry

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Std Var</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogEE</td>
<td>-0.0110</td>
<td>0.0108</td>
<td>-1.02</td>
</tr>
<tr>
<td>LogDay</td>
<td>0.0025</td>
<td>0.0088</td>
<td>0.28</td>
</tr>
<tr>
<td>LogMile</td>
<td>-0.0054</td>
<td>0.0095</td>
<td>-0.57</td>
</tr>
<tr>
<td>LogPotential</td>
<td>-0.2052*</td>
<td>0.0144</td>
<td>-14.23</td>
</tr>
<tr>
<td>Constant</td>
<td>1.2361*</td>
<td>0.1222</td>
<td>10.12</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.4034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Project Cost Distribution</td>
<td>Entry Cost Distribution</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Var.</td>
<td>Mean</td>
</tr>
<tr>
<td>Log(Eng. Est.)</td>
<td>1.1787</td>
<td>0.0427</td>
<td>-0.6899</td>
</tr>
<tr>
<td>Log(Milage)</td>
<td>-0.2107</td>
<td>0.0234</td>
<td>1.7332</td>
</tr>
<tr>
<td>Log(Days)</td>
<td>0.0248</td>
<td>0.0279</td>
<td>-0.5687</td>
</tr>
<tr>
<td>Difference in Const.</td>
<td>-1.5044 (Mean)</td>
<td>0.9251 (Std. Var.)</td>
<td></td>
</tr>
</tbody>
</table>

Table 17: Parameter Estimates for Project Cost and Entry Cost Distributions
APPENDIX D

FIGURES
Figure 1: Estimates of the Inverse Bidding Functions

*: The 1st, 2nd and 3rd rows are estimated by methods 0, 1 and 2.
*: The 1st, 2nd and 3rd columns correspond to \( x = 8, 10, \) and 12.
*: The left and right results are given by methods 0 and 2 respectively.

Figure 2: Estimate of the Conditional Density Function of Private Cost*

Figure 3: Geographical Distribution of Projects and Firms by County
Figure 4: Plot of the Normalized Bids

Figure 5: Plot of Bid vs. Eng. Est.
Figure 6: Histogram of the Normalized Cost Estimates

Figure 8: An Illustrative Example for Model with Bid-preparation Cost
Figure 7: Density Function of the Normalized Cost Estimates
Figure 9: An Illustrative Example for Model with Information-gathering Cost

Figure 10: Histograms of Selected Explanatory Variables in Michigan Market
Figure 11: Distribution of the Regression Errors

Figure 12: Time Series Plot of Latent Variables for a Procurement Auction
Figure 13: Bayesian MCMC Simulation Results (Left: Project Cost Distribution Right: Bid-preparation Cost Distribution)
Figure 14: Comparison between Observed and Simulated Data

Figure 15: Effect of the Number of Potential Bidders on Bid
Figure 16: Effects of N and k on Procurement Cost
BIBLIOGRAPHY


125


[37] Li, T. Perrigne, I. & Young, Quang (2001): "Semiparametric Estimation of the Optimal Reserve Price in First-Price Auctions," working paper, University of Southern California

[38] Li, T. & Zheng X. (2005): "Procurement Auctions with Entry and Uncertain Number of Actual Bidders: Theory, Structural Inference and an Application," working paper. Indiana University


