ANNALS OF ECONOMICS AND FINANCE **25-2**, 675–703 (2024)

A Modified Likelihood Approach for Models with Parameter-Dependent Support

Yao Luo*

In structural models with parameter-dependent support, maximum likelihood problems are nonregular. We reformulate these problems by parameter-dependent transformations and propose modified likelihood estimators that have regular asymptotic properties. We then describe several applications to search models, auction models and frontier production functions and demonstrate the performance of our method through Monte Carlo simulations. Lastly, we apply the method to a Vuong non-nested test of additive v.s. multiplicative separable auction-specific heterogeneity in Michigan Department of Transportation procurements.

Key Words: Boundary model; Job search; Auction; Frontier production function; Sutton's bound test; Vuong test.
JEL Classification Numbers: C13, C12, C57, L11, J64.

1. INTRODUCTION

In structural models, when the decisions of economic agents are endogenized, their supports often change with respect to the parameters. For instance, in auction models, the support of bidders' bids depends on the parameter values even if the support of the value distribution stays the same. This violates the usual regularity conditions of maximum likelihood estimation. Thus, maximum-likelihood estimation of auction models leads to nonregular asymptotics (Donald and Paarsch (1993)). See, e.g., Donald and Paarsch (1993), Hirano and Porter (2003), Chernozhukov and Hong (2004) and Li (2010) for an important literature on studying the properties of the well-known maximum likelihood estimator (MLE) in auctions and alternative estimators. Similar issues also arise in other models due to parameter-dependent boundaries, such as the reserve wage in search mod-

* University of Toronto. Email: yao.luo@utoronto.ca. I thank Gaurab Aryal, Christian Gourieroux, Emmanuel Guerre, Xintong Han, Wei Lin and Ruli Xiao for helpful comments. I acknowledge funding from the SSHRC Insight Grant. The usual disclaimer applies.

1529-7373/2024

els, the maximum output in frontier production functions and the range of prices and qualities offered in nonlinear pricing models.

The aim of this paper is to provide convenient likelihood-based estimators for these models, known as "boundary models". We demonstrate that in many cases it is possible to make use of the usual maximum likelihood estimation procedure with simple modifications. Thus, the contribution of the paper is the convenience of our methods. In particular, we propose a simple transformation to the original maximum likelihood problem in boundary models that leads to a root-N consistent and asymptotically normal MLE. As a result, most practical issues, such as computing an appropriate asymptotic variance, hypothesis testing, specification tests and model selection, have readily available treatments within the framework of maximum likelihood. As an illustration, we apply the proposed method to a Vuong non-nested test of additive v.s. multiplicative separable auction-specific heterogeneity using procurement data from Michigan Department of Transportation.

Consider a random sample $\{x_1, x_2, \ldots, x_N\}$ from a density function $g(x;\theta_0)$, where $\theta_0 \in \Theta \subset R^K$ is an unknown vector of parameters in a given parameter space Θ. The loglikelihood function is defined as

$$L(\theta) = \Sigma_{i=1}^{N} \log g(x_i;\theta).$$

The maximum likelihood estimator, i.e., $\widehat{\vartheta} = \arg\max_{\theta\in\Theta} L(\theta)$, is motivated by the fact that the expected loglikelihood function is maximized at the true parameter value, i.e.

$$E[\log g(x_i;\theta)] \leq E[\log g(x_i;\theta_0)].$$

In boundary models, the support of the random variable $x \sim g(x;\theta)$ depends on the parameter. For simplicity, consider the upper boundary as an example. The support of x is $[0, k(\theta)]$. With an arbitrary θ, the support of the predicted bid distribution may not be covering all the data points, i.e. $k(\theta) < \max_i x_i$, which may lead to undefined loglikelihood due to $g(x_i;\theta) = 0$ for some observation i. One solution is to use a constrained MLE:

$$\widehat{\vartheta} = \arg \max_{\theta: k(\theta) \geq \max_i x_i} \frac{1}{N} \sum_{i=1}^{N} \log g(x_i;\theta),$$

which introduces implementation difficulties. Another major difficulty arises from the complication of the constrained MLE's statistical properties, rendering inference inconvenient.

To avoid such difficulties, we reformulate the maximum likelihood problem. In particular, we consider a new family of density functions, denoted by $\mathrm{g}(\cdot;\theta)$, which we call the induced-likelihood function because it is derived from the original one $g(\cdot;\theta)$. It is also parametrized by θ and satisfies two conditions: (1) $\mathrm{g}(\cdot;\theta_0) = g(\cdot;\theta_0)$; (2) for all $\theta \in \Theta$, $\mathrm{g}(\cdot;\theta)$ has the same support as $g(\cdot;\theta_0)$. Under these two conditions, we can show that the expected likelihood is also maximized at θ_0, i.e. $E[\log \mathrm{g}(x_i;\theta)] \leq E[\log \mathrm{g}(x_i;\theta_0)]$. This motivates our modified likelihood estimator:

$$\widehat{\theta} = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \log \mathrm{g}(x_i;\theta),$$

which involves an unconstrained maximization problem. Moreover, the new estimator has regular asymptotic properties.

Consider again the upper boundary problem. We transform x into a new variable $y = x \cdot k(\theta_0)/k(\theta)$. It is easy to see that the transformed variable y has a density function

$$\mathrm{g}(y;\theta) = \frac{k(\theta)}{k(\theta_0)} \cdot g\big(\frac{k(\theta)}{k(\theta_0)} y;\theta\big),$$

if $y \in [0, k(\theta_0)]$, and $\mathrm{g}(y;\theta) = 0$ otherwise. Note that y has the same support for any parameter value. The next section shows that its expected likelihood maximizes at the true parameter value θ_0. In view of these, we propose a modified likelihood estimator:

$$\widehat{\theta} = \arg\max_{\theta} \frac{1}{N} \sum_{i=1}^{N} \Big[\log k(\theta) + \log g\big(\frac{x_i}{\max_i x_i} k(\theta);\theta\big)\Big],$$

where $k(\theta_0)$ is replaced by $\max_i x_i$ to make the estimator feasible. The first term in the brackets reflects a contribution of the varying boundary to the likelihood. Moreover, the varying boundary is used to transform all data points into the support of the predict random variable x at this parameter value θ. Thus, the terms that we apply logarithm on are never zero.

Our modified likelihood approach has many applications in structural models with parameter-dependent supports. In job search models, Flinn and Heckman (1982) use the maximum likelihood method under the constraint that the predicted reserve wage equals the observed minimum wage. This leads to a constrained maximization problem and hence introduces implementation difficulties. On the contrary, our modified MLE involves an unconstrained maximization problem. Assuming a special parameter that is monotone in the bound of the support, Christensen and Kiefer (1991) proposed a profile likelihood method. They use a super consistent estimate of the boundary to concentrate out the special parameter and then estimate the rest of the parameters via maximum likelihood. While our approach also contains a first-step estimation of the boundaries, it does not require a special parameter.

Another important application of our method is structural analysis of auction data. See, e.g., Paarsch (1992), Guerre, Perrigne, and Vuong (2000) and Athey and Haile (2007). Our approach complements existing parametric approaches.[1] Donald and Paarsch (1993) propose pseudo-maximum likelihood estimation. Similar to the above-mentioned profile likelihood approach, their idea is to concentrate out one parameter using a super-consistent estimate of boundary. However, this requires all the exogenous variables being discrete. On the other hand, simulation-based methods provide root-N asymptotically normal estimates in general. Laffont and Vuong (1993) suggest simulated nonlinear least squares and simulated method of moment for estimating descending price auctions. Li (2010) and Li and Zhang (2015) propose the indirect inference approach for estimating symmetric and asymmetric first-price auctions, respectively. Recently, Aryal, Gabrielli, and Vuong (2021) propose the method of moments using private values that are estimated using local polynomial methods.

The literature of parametric deterministic frontier function considers a linear model where the error has a one-sided distribution such as exponential or half-normal.[2] See Amsler, Leonard, and Schmidt (2013) for a brief review of estimation and inference in deterministic frontier models. Schmidt (1976) showed that the Aigner and Chu (1968) linear programming estimator was the MLE under certain parametric assumptions. However, he noted that the statistical properties were unknown due to violations of the usual regularity conditions. Under the assumption that the density of the error and its derivative are both zero at the boundary, Greene (1980) showed that the problem becomes regular. When this assumption fails, our approach provides a three-step estimator that has regular properties, which complements Greene (1980).

Last but not least, there are papers that provide general treatments to likelihood-based estimation in boundary models. Hirano and Porter (2003) show that the maximum likelihood estimator is generally inefficient, but that the Bayes estimator is efficient according to the local asymptotic minmax criterion for conventional loss functions. Chernozhukov and Hong (2004) propose likelihood-based estimation and inference methods for the general class of structural models with a jump in the conditional density. By focusing optimal (Bayes) and maximum likelihood procedures, they derive convergence rates and distribution theory. Despite its poten-

[1]Guerre, Perrigne, and Vuong (2000) propose a general nonparametric approach that is based on the bidder's first-order condition.

[2]Similar applications include Sutton's bound test (Sutton (1991) and Bronnenberg, Dhar, and Dubé (2011)) and auction models with separable auction-specific covariates (Haile, Hong, and Shum (2003)).

tial inefficiency relative to the Bayes one, our approach remains within the likelihood framework but circumvents the difficulties of the classical MLE arising from parameter-dependent support.

The rest of the paper is structured as follows. Section 2 introduces the main idea and several methods of transformation. Section 3 describes three applications in economics: search models, auction models and frontier production functions. Section 4 contains some Monte Carlo experiments, while Section 5 illustrates an empirical application to DOT procurement auctions. Section 6 concludes.

2. MAIN IDEA

In this section, we consider the general parametric maximum likelihood problems without covariates. We are interested in estimating the distribution of a random variable x with unknown support $[k_L(\theta), k_R(\theta)]$. Denote the distribution and density functions as $G(\cdot;\theta)$ and $g(\cdot;\theta)$, respectively. The true value of the parameter is denoted by θ_0 and the support $[k_L(\theta_0), k_R(\theta_0)]$. Without loss of generality, hereafter, we assume that the original loglikelihood function identifies the true parameter.

Consider a vector of population statistics μ. We define a transformed variable

$$y = T(x;\mu,\theta),$$

such that the following two conditions are satisfies:

(1) $\mathrm{g}(\cdot;\theta_0) = g(\cdot;\theta_0)$, and

(2) y has the support $[k_L(\theta_0), k_R(\theta_0)]$,

where $\mathrm{g}(\cdot;\theta)$ represents the density function of y.[3] For example, in the upper boundary problem, we let $\mu = k_R(\theta_0)$ and $T(x;\mu,\theta) = x\cdot k_R(\theta_0)/k_R(\theta)$.

Under these two conditions, we have

$$\begin{aligned} E[\log\frac{\mathrm{g}(x;\theta)}{\mathrm{g}(x;\theta_0)}] < \log E[\frac{\mathrm{g}(x;\theta)}{\mathrm{g}(x;\theta_0)}] &= \log\int_{k_L(\theta_0)}^{k_R(\theta_0)}\frac{\mathrm{g}(x;\theta)}{\mathrm{g}(x;\theta_0)}g(x;\theta_0)dx \\ &= \log\int_{k_L(\theta_0)}^{k_R(\theta_0)}\frac{\mathrm{g}(x;\theta)}{\mathrm{g}(x;\theta_0)}\mathrm{g}(x;\theta_0)dx = 0, \end{aligned}$$

where we use the facts that $\mathrm{g}(\cdot;\theta_0) = g(\cdot;\theta_0)$. Therefore, we can transform the original problem into a new maximum likelihood problem as summarized in the following Theorem.

[3]There exists at least a linear transformation such that conditions (1) and (2) are satisfied:

$$y = \frac{k_L(\theta_0)k_R(\theta) - k_R(\theta_0)k_L(\theta)}{k_R(\theta) - k_L(\theta)} + \frac{k_R(\theta_0) - k_L(\theta_0)}{k_R(\theta) - k_L(\theta)}x,$$

where the vector of population statistics is $\mu = (k_L(\theta_0), k_R(\theta_0))'$.

THEOREM 1. *The expectation of y's loglikelihood function is maximized at θ_0, i.e.*

$$E[\log \mathrm{g}(x;\theta)] \leq E[\log \mathrm{g}(x;\theta_0)].$$

Theorem 1 says that the expected loglikelihood is maximized at the true parameter value. On the other hand, it is silent on whether we can identify the same vector of parameters or a subvector of them. In later sections, we will see a few examples in which the transformed likelihood only identifies a subvector of the parameters.

2.1. Methods of Transformation

In this section, we discuss a few methods of transformation to generate new families of likelihood function $\mathrm{g}(\cdot;\theta)$ that satisfy the above two conditions. The modified likelihood methods differ by: (1) the population statistics μ that we obtain from data, e.g. minimum and maximum; (2) the function form of the transformation $T(\cdot)$, e.g. scaling, shifting and power.

2.1.1. Scaling

The scaling transformation defines a new variable $y = x \cdot c(\theta)$, where $c(\theta_0) = 1$.

For example, the Pareto distribution has a density function

$$g(x;\theta) = \frac{\theta_2 \theta_1^{\theta_2}}{x^{\theta_2+1}},$$

where $\theta_1, \theta_2 > 0$ and $x \geq \theta_1$. Therefore, the likelihood problem here involves a parameter-dependent support $[\theta_1, +\infty]$. We have a left boundary problem where $k_L(\theta) = \theta_1$.

We consider a transformed variable $y = x \cdot k_L(\theta_0)/\theta_1$, which has a density function

$$\mathrm{g}(y;\theta) = \frac{\theta_1}{k_L(\theta_0)} \cdot g(\frac{\theta_1}{k_L(\theta_0)} y;\theta) = \frac{\theta_2}{k_L(\theta_0)} \frac{k_L(\theta_0)^{\theta_2+1}}{y^{\theta_2+1}},$$

and the expected loglikelihood function

$$\log \theta_2 - (\theta_2 + 1)(\log y - \log k_L(\theta_0)).$$

This gives the maximum likelihood estimator for θ_2

$$\widehat{\theta}_2 = \frac{1}{\frac{1}{N}\sum_{i=1}^N \log x_i - \log \min_i x_i},$$

where $k_L(\theta_0)$ has been replaced by $\min_i x_i$. Note that the new likelihood function only identifies θ_2. The scaling transformation separates the parameter vector θ into: (1) θ_1 as the location parameter and (2) θ_2 as the shape parameter.

2.1.2. Shifting

The shifting transformation defines a new variable $y = x + c(\theta)$, where $c(\theta_0) = 0$.

Consider again the Pareto distribution. The transformed variable is $y = x + (k_L(\theta_0) - \theta_1)$, which has a density function

$$\mathrm{g}(y;\theta) = g(y + \theta_1 - k_L(\theta_0)),$$

which implies the expected loglikelihood function

$$\log\theta_2 + \theta_2\log\theta_1 - (\theta_2+1)E[\log(x - k_L(\theta_0) + \theta_1)].$$

This suggests a modified maximum likelihood estimator for θ

$$\widehat{\theta} = \arg\max_{\theta}\Big\{\log\theta_2 + \theta_2\log\theta_1 - \frac{1}{N}(\theta_2+1)\sum_{i=1}^{N}[\log(x_i - \min_i x_i + \theta_1)]\Big\},$$

where $k_L(\theta_0)$ has been replaced by $\min_i x_i$. Note that the shifting transformation keeps both θ_1 and θ_2 as shape parameters.

2.1.3. Power

The power transformation defines a new variable $y = x^{c(\theta)}$, where $c(\theta_0) = 1$.

Example: suppose x is uniform distributed on $[0, \theta]$.[4]

While a scaling transformation seems a natural choice for the uniform distribution on $[0, \theta]$, it does not work. The reason is that the only parameter here is the scale parameter. In fact, the original density is $g(x;\theta) = 1/\theta$ for $x \in [0, \theta]$. The transformed variable $y = x \cdot k_R(\theta_0)/\theta$ has a density $\mathrm{g}(y;\theta) = 1/k_R(\theta_0)$. In other words, the density of the transformed variable becomes flat or not informative of the parameter after our transformation, i.e. $E[\log\frac{\mathrm{g}(x;\theta)}{\mathrm{g}(x;\theta_0)}] = 1, \forall\theta$.

Instead, we consider a power transformation. In the first step, we add one unit to every data point so that we can consider estimating θ for a

[4]This example is for demonstration purposes only. A more obvious estimator is $2\sum_{i=1}^{N} x_i/N$, which is root-N consistent and asymptotically normal.

uniform distribution $[1, 1+\theta]$ without loss of generality. In the second step, we consider the transformation variable $y = x^{\log k_R(\theta_0)/\log(1+\theta)}$. Note that $\log k_R(\theta_0), \log(1+\theta) > 0$. Moreover, y has a distribution function

$$\begin{aligned} \mathbb{G}(y;\theta) &= \Pr[x^{\log k_R(\theta_0)/\log(1+\theta)} \leq y] = \Pr[\log x \leq \frac{\log y \log(1+\theta)}{\log k_R(\theta_0)}] \\ &= \frac{\exp[\frac{\log y \log(1+\theta)}{\log k_R(\theta_0)}] - 1}{\theta}, \end{aligned}$$

and a density function

$$\mathbb{g}(y;\theta) = \frac{\exp[\frac{\log y \log(1+\theta)}{\log k_R(\theta_0)}]}{\theta} \frac{\log(1+\theta)}{\log k_R(\theta_0)} \frac{1}{y},$$

which leads to expected loglikelihood function

$$-\log\theta + \log\log(1+\theta) + \frac{E[\log y]\log(1+\theta)}{\log k_R(\theta_0)} - E[\log y],$$

and a modified likelihood estimator

$$\widehat{\theta} = \arg\max_{\theta} \Big\{ -\log\theta + \log\log(1+\theta) + \frac{1}{N}\sum_{i=1}^{N} \log x_i \cdot \Big(\frac{\log(1+\theta)}{\log \max_i x_i} - 1\Big)\Big\},$$

where $k_R(\theta_0)$ has been replaced by $\max_i x_i$. This transformation turns a location parameter of the uniform distribution into a shape parameter of a new distribution.

2.2. Discussion

It is well known that the likelihood is invariant to transformations of data when the transformations are parameter independent. That is, the likelihood of $\{y_i = T(x_i)\}$ is

$$\Pi_{i=1}^{N} g(T^{-1}(y_i);\theta)\frac{\partial T^{-1}(y)}{\partial y}|_{y=y_i} = \Pi_{i=1}^{N} g(x_i;\theta)\frac{\partial T^{-1}(y)}{\partial y}|_{y=y_i},$$

which is proportional to the likelihood of $\{x_i\}$. Our transformations are different because they depend on the parameter θ so as to match the predicted support with the true one. For the same reason, our proposed induced-likelihood function is different from marginal likelihood, conditional likelihood and partial likelihood.

Moreover, our modified likelihood methods are also different from the profiled likelihood method of Christensen and Kiefer (1991) and Donald and Paarsch (1993). Take the latter as an example. In first-price auctions, since the lower boundary of the bid equals the lower boundary of the valuation, one could have a super consistent estimate of the latter using the former. Donald and Paarsch (1993) use this estimate to concentrate out a parameter and the maximum likelihood problem for the rest of the parameters becomes regular. Under the assumption that the lower bound $k_L = k_L(\theta_1, \theta_2)$ is monotone in θ_1, their pseudo-MLE solves the following problem:

$$\max_{\theta_2} \sum_{i=1}^{N} \log g\Big(b_i; \theta_1(\theta_2, \widehat{k}_L), \theta_2\Big),$$

where $\theta_1(\theta_2, k_L)$ is the inversion of the mapping from the parameter to the lower bound $k_L = k_L(\theta_1, \theta_2)$, and $\widehat{k}_L \equiv \min_i b_i$ represents the minimum bid. Obviously, this method is inapplicable to models with a single parameter, such as auction models with power distributed valuations.

On the other hand, our method does not require a special parameter that is monotone in the bound of the support. Instead of adjusting the parameter vector to "profile" the likelihood function, we adjust the bids to fit them in the predicted support of the bid distribution $[0, k(\theta)]$. This adjustment adds an extra term in the loglikelihood function to reflect the change of support. Moreover, the method extends to the case when lower bound changes with the parameters, for which we consider the transformed variable $y = x \cdot k(\theta_0)/k(\theta)$, where $k(\theta)$ is the lower bound of the variable x when the parameter value is θ, and $x \in [k(\theta), +\infty]$.[5]

3. APPLICATIONS

In this section, we describe a few applications of our method in economics.

[5]It is interesting to point out that our estimator coincides with the pseudo MLE of Donald and Paarsch (1993) in the Pareto case (see their subsection 4.2 on page 132). In fact, $k(\theta) = \frac{\theta_1 \theta_2 I}{\theta_2 I - 1}$, and

$$\log \mathrm{g}(y; \theta) = \log k(\theta) + [\log(\theta_2(I+1)) + \theta_2(I+1)\log k(\theta)] - [\theta_2(I+1)+1]\log[k(\theta)y/\underline{b}].$$

Maximizing its empirical counterpart leads to the transformed MLE:

$$\widehat{\theta}_2 = \frac{1}{(I+1)(\frac{1}{N}\sum_{i=1}^{N} \log b_i - \log \underline{b})}.$$

3.1. Search Models

In this subsection, we consider a simplified version of a search model. In discrete time, workers live infinitely and discount future payoffs at rate β. At each period, a worker can be either employed or unemployed. An unemployed worker receives a wage offer, which is a random draw from the density $f(\cdot;\theta)$. If a job offer is accepted, the worker is employed for forever. It is easy to see that the worker's problem has a cutoff property: there exists a cutoff value ξ such that the worker accept an offer if and only if it is higher than ξ. Assume that the wage offer density function $f(\cdot;\theta)$ is time-invariant.

Denote the distribution of wage offers as $F(\cdot;\theta)$. The probability of observing a person unemployed for t_i periods and then accepted a wage offer w_i is

$$F(\xi;\theta)^{t_i} \cdot [1 - F(\xi;\theta)] \cdot \frac{f(w_i;\theta)}{1 - F(\xi;\theta)} \cdot 1(w_i \geq \xi)$$

where $F(\xi;\theta)^{t_i} \cdot [1 - F(\xi;\theta)]$ is the probability that the person stays unemployed for t_i periods and then accepts an offer, and $\frac{f(w_i;\theta)}{1-F(\xi;\theta)}$ is the conditional density of his/her wage conditional on $w_i \geq \xi$. The density function is

$$g(w,t;\theta,\xi) = F(\xi;\theta)^{t_i} f(w_i;\theta) 1(w_i \geq \xi)$$

which has a support $[\xi, +\infty)$. To estimate θ, Christensen and Kiefer (1991) proposed a profile likelihood method that requires the boundary is monotone in a special parameter. A similar approach was adopted in some auction papers such as Donald and Paarsch (1993).

Consider now our modified likelihood methods. Consider a shifting transformation and the resulting transformed variable $y = w + (\underline{w} - \xi)$. y has a density function

$$\mathrm{g}(y,t;\theta,\xi) = F(\xi;\theta)^{t_i} f(y - \underline{w} + \xi;\theta),$$

where $\underline{w}$ is the minimum wage offer that has been accepted. This leads to a modified likelihood estimator:

$$\widehat{\theta} = \arg\max_{(\theta,\xi)} \frac{1}{N} \sum_{i=1}^{N} \Big[t_i \log F(\xi;\theta) + \log f(w_i - \widehat{\underline{w}} + \xi;\theta) \Big],$$

where $\widehat{\underline{w}} = \min_i w_i$ is the minimum accepted wage offer.

On the other hand, one can consider a scaling transformation and the resulting transformed variable $y = x\underline{w}/\xi$, whose support is $[\underline{w}, +\infty)$. Its density function is

$$g(y;\theta) = \frac{\xi}{\underline{w}} \cdot F(\xi;\theta)^t \cdot f(\frac{\xi}{\underline{w}} y;\theta).$$

Thus, the transformed MLE can be defined similarly:

$$\widehat{\theta} = \arg\max_{(\theta,\xi)} \frac{1}{N} \sum_{i=1}^{N} \Big[t_i \log F(\xi;\theta) + \log f(w_i \cdot \xi/\widehat{\underline{w}};\theta) + \log \xi \Big].$$

Normal Distribution

Consider the normal distribution with mean μ and standard deviation σ as the wage distribution. Under a shifting transformation, y has a density function

$$g(y;\theta) = \Phi\Big(\frac{\xi-\mu}{\sigma}\Big)^t \cdot \phi\Big(\frac{(y-\underline{w}+\xi)-\mu}{\sigma}\Big)/\sigma,$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution function and the density function of the standard normal variable, respectively. It is obvious that ξ and μ are not separately identified through $g(\cdot;\theta)$ but their difference $\xi - \mu$ is. Thus, we propose the following estimator for (μ, σ):

$$\widehat{\theta} = \arg\max_{(\mu,\sigma)} \frac{1}{N} \sum_{i=1}^{N} \Big[t_i \log \Phi(\frac{\widehat{\xi}-\mu}{\sigma}) + \log \phi(\frac{w_i-\mu}{\sigma}) - \log \sigma \Big],$$

where $\widehat{\xi} = \min_i w_i$.

A scaling transformation leads to a density function

$$g(y;\theta) = \frac{\xi/\sigma}{\underline{w}} \cdot \Phi\Big(\frac{\xi-\mu}{\sigma}\Big)^t \cdot \phi\Big(\frac{y\cdot\xi/\underline{w}-\mu}{\sigma}\Big).$$

It is obvious that the three parameters are not identified separately through $g(\cdot;\theta)$ but ξ/σ and μ/σ are. After some algebra, we obtain the same estimator for (μ, σ) as the one obtained under a shifting transformation. Standard derivations give the asymptotic distribution of $\widehat{\theta}$:

$$\sqrt{N}(\widehat{\theta} - \theta) \overset{A}{\sim} N(0, \Sigma),$$

where Σ represents the asymptotic variance.

3.2. Auction Models

In this subsection, we introduce the first-price auction model. Consider independent private value first-price auctions where I bidders are symmetric and risk neutral. Their private values are *i.i.d.* draws from a common distribution $F(\cdot;\theta)$, which is determined by a (finite-dimensional or infinite-dimensional) parameter θ. This distribution is absolutely continuous with density $f(\cdot;\theta)$. Assuming that there is no reserve price, a bidder with value v solves the following problem:

$$\max_b (v-b)F(s^{-1}(b;\theta);\theta)^{I-1},$$

where $s^{-1}(\cdot;\theta)$ is the inverse bidding strategy, $(v-b)$ is the profit if he/she wins, and $F(s^{-1}(b;\theta);\theta)^{I-1}$ is his/her probability of winning if he/she bids b. The equilibrium bidding strategy is

$$b = s(v;\theta) \equiv v - \frac{1}{F(v;\theta)^{I-1}} \int_0^v F(x;\theta)^{I-1}dx,$$

which is increasing with respect to the value v.

Power distribution with known support $[0,1]$

Consider independent private value auctions with zero reserve price and risk neutral bidders. The value distribution is $F(v) = v^\theta$, where $v \in [0,1]$. The bidding strategy is $b(v) = k(\theta)v$, where $k(\theta) = [1 - \frac{1}{\theta(I-1)+1}]$. Thus, the bid has a support of $[0, k(\theta)]$. The bid distribution function is $G(b;\theta) = (b/k(\theta))^\theta$ and its density function is $g(b;\theta) = \theta(b/k(\theta))^{\theta-1}/k(\theta)$, where $b \in [0, k(\theta)]$. Note that even though the support of the value distribution stays the same $[0,1]$, the upper bound of the bid distribution $k(\theta)$ depends on the parameter θ.

Consider reformulating the maximum likelihood problem. Define a transformed variable as

$$y \equiv x\bar{b}/k(\theta),$$

where $x \in [0, k(\theta)]$. The transformed variable has a density function $\mathrm{g}(y;\theta) = \theta(y/\bar{b})^{\theta-1}/\bar{b}$, where $y \in [0,\bar{b}] \equiv [0, k(\theta_0)]$. Now consider the following expected loglikelihood function:

$$\begin{aligned} E[\log \mathrm{g}(b;\theta)] &= \log\theta + (\theta-1)\{E[\log b] - \log\bar{b}\} - \log\bar{b} \\ &= \log\theta - \log\bar{b} - (\theta-1)/\theta_0, \end{aligned}$$

which is maximized at $\theta = \theta_0$. The expectation is taken with respect to the true random variable b and we have replaced y with b in the loglikelihood function. Note that y has the same support for all parameter values and its expected likelihood function is maximized at the true parameter value θ_0.

Maximizing the empirical counterpart of $E[\log \mathrm{g}(b;\theta)]$, i.e.,

$$\max_{\theta} \log \theta + (\theta - 1)((\sum_{i=1}^{N} \log b_i)/N - \log \max_i b_i),$$

leads to our new MLE

$$\widehat{\theta} = \frac{1}{\log(\max_i b_i) - (\sum_{i=1}^{N} \log b_i)/N},$$

where we have replaced $\bar{b}$ by $\max_i b_i$.[6] Note that the asymptotic distribution of $\widehat{\theta}$ does not depend on that of the superconsistent estimator of the upper bound $\max_i b_i$. It is easy to see that this new estimator is regular. Thus, the usual asymptotic results apply.

We also remark that the first-order derivative of $E[\log \mathrm{g}(b;\theta)]$ is $1/\theta - 1/\theta_0$ and the second-order derivative is $-1/\theta^2$. So the Fisher information is $I(\theta) = -E[-1/\theta^2] = 1/\theta^2$ and the asymptotic variance of our transformed MLE is $1/I(\theta_0) = \theta_0^2$. In sum, the new MLE is root-N consistent and asymptotically normal:

$$\sqrt{N}(\widehat{\theta} - \theta_0) \overset{A}{\sim} N(0, \theta_0^2).$$

Power distribution with unknown support $[0, m]$

In empirical applications, it is unusual that the analyst knows ex ante the support of the value distribution. So it is desirable to develop a feasible estimator for θ when the support is unknown. Assume that the value distribution is $F(v) = (v/m)^\theta$, where $v \in [0, m]$.[7] Following the same lines, we can show that the bidding strategy is $b(v) = k(\theta)v$, the bid distribution is $G(b;\theta) = (\frac{b}{mk(\theta)})^\theta$, $g(b;\theta) = \theta(\frac{b}{mk(\theta)})^{\theta-1}/(mk(\theta))$, where $b \in [0, mk(\theta)]$.

[6]Alternatively, we can exploit the fact that $E[\log b] = \log \bar{b} - 1/\theta_0$, which leads to the same estimator $\widehat{\theta} = 1/[\log(\max_i b_i) - (\sum_{i=1}^{N} \log b_i)/N]$.

[7]If the value distribution has a unknown support, without loss of generality, assume the valuation is generated by multiplying a valuation on $[0, 1]$ by a scale parameter $m > 0$. It is well known that this multiplicative separability passes to the bid distribution as well. See Krasnokutskaya (2011). So the new generic variable becomes $\widetilde{x} = mx$. Consider the new transformed variable $\widetilde{y} = \widetilde{x}\bar{b}/(mk(\theta)) = x\bar{b}/k(\theta) = y$, where $\widetilde{x} \in [0, mk(\theta)]$. Therefore, $\mathrm{g} = \widetilde{\mathrm{g}}$ and $E[\log \mathrm{g}(b;\theta)] = E[\log \widetilde{\mathrm{g}}(b;\theta)]$.

The transformed variable has a density function $\mathrm{g}(y;\theta) = \theta(y/\bar{b})^{\theta-1}/\bar{b}$, where $y \in [0,\bar{b}] \equiv [0, m_0 k(\theta_0)]$. Thus, the expected loglikelihood function

$$E[\log \mathrm{g}(b;\theta)] = \log\theta + (\theta-1)\{E[\log b] - \log\bar{b}\} - \log\bar{b} = \log\theta - \log\bar{b} - (\theta-1)/\theta_0,$$

where $E[\log b] = \log\bar{b} - 1/\theta_0$. Obviously, it is maximized at $\theta = \theta_0$. Notice that this function does not contain the scale parameter m.

Maximizing the empirical counterpart of $E[\log \mathrm{g}(b;\theta)]$ leads to the same estimator for θ. Moreover, we estimate the scale parameter m by

$$\widehat{m} = \frac{\max_i b_i}{k(\widehat{\theta})} = \frac{\max_i b_i}{1 - \frac{1}{\widehat{\theta}(I-1)+1}}.$$

which is a multiplier that matches the observed maximum bid with the predicted upper bound $k(\widehat{\theta})$ at the estimated parameter value $\widehat{\theta}$. Since $\max_i b_i$ is superconsistent and $\widehat{m}$ is a smooth function of $\widehat{\theta}$, the estimator for the scale parameter also has regular asymptotics. In particular,

$$\sqrt{N}(\widehat{m} - m_0) \overset{A}{\sim} N(0, \Big[\frac{m_0}{\theta_0(I-1)+1}\Big]^2).$$

3.3. Frontier Production Functions

In this subsection, we apply our method to estimation of parametric frontier production function.[8] We follow closely Schmidt (1976) for the model part. Consider a Cobb-Douglas frontier production function in log form:

$$x_i = \beta_0 + z_i'\beta_1 - \epsilon_i,$$

where x_i is the log of plant i's output, given the logs of the inputs $z_i = (z_{i1}, \ldots, z_{iK})'$. Since we interpret the "fitted value" $\overline{x}(z_i) = \beta_0 + z_i'\beta_1$ as the maximum output, the disturbance ϵ_i is assumed to be positive. Thus, the random error represents the factors that result in less than maximum output. Thanks to our Cobb-Douglas structure, the total of all elements of $\beta_1 = (\beta_{11}, \ldots, \beta_{1K})'$, i.e., $\sum_{t=1}^{K} \beta_{1t}$, conveniently represents the return to scale.

Following Smith (1985), we make the following assumption on the density function of the disturbance ϵ, $f(\cdot)$.

[8]We remark that this method can be applied to auctions with covariates (Guerre, Perrigne, Vuong (2000) and Haile, Hong, and Shum (2003)) and Sutton's bound test (Sutton (1991) and Bronnenberg, Dhar, and Dubé (2011)).

Assumption 1. *$f(\epsilon) \sim \alpha\gamma\epsilon^{\alpha-1}$ as $\epsilon \downarrow 0$, ($\alpha > 0, \gamma > 0$).*

This assumption, known as Paretian tail decay, is quite general. It allows three-parameter Weibull, three-parameter gamma, three-parameter beta and three-parameter log gamma. Note that the commonly assumed distributions, such as exponential and half-normal, satisfy this assumption. For $\alpha > 2$, the usual maximum likelihood estimates are consistent, asymptotically efficient and asymptotically normal. See Greene (1980). However, these properties are no longer valid for $0 < \alpha \leq 2$. We seek to provide an estimator with such properties when $0 < \alpha \leq 2$. In particular, for $0 < \alpha \leq 2$, our modified likelihood approach leads to a three-step estimator that is root-N consistent and asymptotically normal. We use the two commonly adopted distributions to illustrate the three steps.[9]

Exponential Distribution

Suppose that the error ϵ_i follows an exponential distribution

$$f(\epsilon) = \gamma \exp(-\gamma\epsilon),$$

where $\epsilon > 0$. Note that the exponential distribution is a special case of the gamma distribution. The mean of this distribution is $1/\gamma$. Given the inputs z, the conditional density function of the output x is

$$g(x|z;(\beta,\gamma)) = \gamma \exp\Big(-\gamma(\beta_0 + z'\beta_1 - x)\Big).$$

In the first step, we obtain an estimate of the population statistic $\overline{x}(z)$. Just for this step, following Smith (1994), we normalize the sum of the z_i to the zero vector for the estimation procedure, without loss of generality. Exploiting the fact that the observed output is smaller than the maximum one, i.e., $x_i < \beta_0 + z_i'\beta_1$, we then propose a first-step estimator:

$$\widetilde{\beta} = \arg \max_{\beta: x_i < \beta_0 + z_i'\beta_1} \beta_0,$$

which is the linear programming estimator proposed by Aigner and Cuh (1968). See Smith (1994) for asymptotic results for the linear programming estimator $\widetilde{\beta}$. It is worth mentioning that if ϵ_i are indepedent with common

[9]Economic theory may provide predictions on the associated tail index α, as well. For instance, in first-price auctions with a binding reserve price, Hill and Shneyerov (2013) find that $\alpha = 1/2$ under private value, and $= 1$ under common value.

density function that converges to $\alpha\gamma\epsilon^{\alpha-1}$ as $\epsilon \downarrow 0$, the linear programming estimator is consistent at rate $O(N^{1/\alpha})$. In our case, the common density function $f(\cdot)$ converges to $\gamma > 0$ as $\epsilon \downarrow 0$, which implies that $\alpha = 1$. Therefore, our first step estimator is consistent at rate $O(N)$.[10] Unfortunately, its actual asymptotic distribution is complicated. In sum, the first step mainly provides a superconsistent estimate for the maximum output $\widehat{\overline{x}}(z) = \widetilde{\beta}_0 + z'\widetilde{\beta}_1$.

In the second step, we propose a modified likelihood estimator for γ based on the linear programming estimator $\widetilde{\beta}$. In particular, we consider a shifting transformation $y_1 = x - (\beta_0 + z'\beta_1) + \overline{x}(z)$. The density of y_1 becomes

$$\mathrm{g}_1(y_1|z;(\beta,\gamma)) = g(y_1 + (\beta_0 + z'\beta_1) - \overline{x}(z)|z;(\beta,\gamma)),$$

which leads to the induced loglikelihood function

$$\begin{aligned} \log \mathrm{g}_1(y_1|z;(\beta,\gamma)) &= \log\gamma - \gamma\left[(\beta_0 + z'\beta_1) - \Big(y_1 + (\beta_0 + z'\beta_1) - \overline{x}(z)\Big)\right] \\ &= \log\gamma - \gamma[\overline{x}(z) - y_1]. \end{aligned}$$

Thus, we can defined our modified likelihood estimator

$$\widehat{\gamma} = \arg\max_{\gamma}\Big\{\log\gamma - \gamma\cdot\frac{1}{N}\sum_{i=1}^{N}[\widehat{\overline{x}}(z_i) - x_i]\Big\} = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}[\widetilde{\beta}_0 + z_i'\widetilde{\beta}_1 - x_i]},$$

which is exactly the same as the one in Schmidt (1976). It is easy to see that

$$\sqrt{N}(\widehat{\gamma} - \gamma) \overset{A}{\sim} N(0, 1/\gamma^2),$$

where γ represents the true parameter value.

In the third step, we propose an estimator for β by considering a scaling transformation $y_2 = x\cdot\overline{x}(z)/(\beta_0 + z'\beta_1)$. The density of y_2 becomes

$$\mathrm{g}_2(y_2|z;(\beta,\gamma)) = g\big(y_2\cdot(\beta_0 + z'\beta_1)/\overline{x}(z)|z;(\beta,\gamma)\big)\cdot(\beta_0 + z'\beta_1)/\overline{x}(z),$$

which leads to a modified likelihood estimator for β:

$$\widehat{\beta} = \arg\max_{\beta}\frac{1}{N}\sum_{i=1}^{N}\Big[\log\big(\widehat{\gamma}(\beta_0 + z_i'\beta_1)\big) - \widehat{\gamma}(1 - \frac{x_i}{\widehat{\overline{x}}(z_i)})\cdot(\beta_0 + z_i'\beta_1)\Big].$$

[10]See subsection A.1 for Monte Carlo experiments that confirm this claim.

It worths mentioning that γ is replaced by the second step estimator $\widehat{\gamma}$ because we can only identify $(\gamma\beta_0, \gamma\beta_1)$ through $\mathrm{g}_2(\cdot)$.

Note that the first step estimator is consistent at rate $O(N)$. See Smith (1994) for details. Thus, it does not affect the asymptotic distributions of the later two estimators $\widehat{\gamma}$ and $\widehat{\beta}$. This renders our problem similar to a standard two-step maximum likelihood estimation. Following Murphy and Topel (2002), we obtain

$$\sqrt{N}(\widehat{\beta} - \beta) \overset{A}{\sim} N(0, \Sigma),$$

where β represents the true parameter vector,

$$\Sigma = R_2^{-1} + R_2^{-1}[R_3' R_1^{-1} R_3 - R_4' R_1^{-1} R_3 - R_3' R_1^{-1} R_4] R_2^{-1},$$

and

$$R_1 = -E\left[\frac{\partial^2 \log \mathrm{g}_1}{\partial\gamma\partial\gamma'}\right], \qquad R_2 = -E\left[\frac{\partial^2 \log \mathrm{g}_2}{\partial\beta\partial\beta'}\right],$$
$$R_3 = -E\left[\frac{\partial^2 \log \mathrm{g}_2}{\partial\gamma\partial\beta'}\right], \qquad R_4 = E\left[\frac{\partial \log \mathrm{g}_1}{\partial\gamma}\frac{\partial \log \mathrm{g}_2}{\partial\beta'}\right].$$

Half-Normal Distribution

Another example is half-normal distributions whose density functions

$$f(\epsilon) = \frac{2}{\sqrt{2\pi}\sigma}\exp(-\frac{\epsilon^2}{2\sigma^2})$$

if $\epsilon \geq 0$, and $f(\epsilon) = 0$ otherwise. Note that the half-normal distribution is a special case of the gamma distribution. Thus, the conditional density function of output is

$$g(x|z; (\beta, \sigma)) = \frac{2}{\sqrt{2\pi}\sigma}\exp(-\frac{(\beta_0 + z'\beta_1 - x)^2}{2\sigma^2}).$$

In the first step, we can use the same linear programming estimator $\widetilde{\beta}$ and obtain a preliminary estimate of the maximum output $\widehat{\overline{x}}(z) = \widetilde{\beta}_0 + z'\widetilde{\beta}_1$. In this case, the common density function $f(\cdot)$ converges to $\frac{2}{\sqrt{2\pi}\sigma} > 0$ as $\epsilon \downarrow 0$, which implies that $\alpha = 1$. Therefore, our first step estimator is consistent at rate $O(N)$.

In the second step, a shifting transformation leads to a density function

$$\mathrm{g}_1(y_1|z; (\beta, \sigma)) = g(y_1 + (\beta_0 + z'\beta_1) - \overline{x}(z)|z; (\beta, \sigma)) = \frac{2}{\sqrt{2\pi}\sigma}\exp(-\frac{(\overline{x}(z) - y_1)^2}{2\sigma^2}).$$

After some algebra, we obtain a simple modified likelihood estimator for σ

$$\widehat{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[(\widehat{\overline{x}}(z_i) - x_i)^2]}.$$

In the third step, a scaling transformation leads to a density function

$$\mathrm{g}_2(y_2|z;(\beta,\sigma)) = g\big(y_2 \cdot \frac{\beta_0 + z'\beta_1}{\overline{x}(z)}|z;(\beta,\gamma)\big)\frac{\beta_0 + z'\beta_1}{\overline{x}(z)}.$$

After some algebra, we obtain a modified likelihood estimator for β

$$\widehat{\beta} = \arg\max_{\beta} \frac{1}{N}\sum_{i=1}^{N}\Big[\log(\beta_0 + z_i'\beta_1) - \frac{1}{2\widehat{\sigma}^2}[(\beta_0 + z_i'\beta_1)\cdot(1 - \frac{x_i}{\widehat{\overline{x}}(z_i)})]^2\Big].$$

Again, σ is replaced by the second step estimator $\widehat{\sigma}$ because we can only identify $(\sigma\beta_0, \sigma\beta_1)$ through $\mathrm{g}_2(\cdot)$. The asymptotic distributions of $\widehat{\sigma}$ and $\widehat{\beta}$ can be derived similarly.

4. MONTE CARLO EXPERIMENTS

To demonstrate the performance of our proposed methods, we run a list of Monte Carlo experiments using the above mentioned structural models with parameter-dependent support.

4.1. Search Models

Consider the normal distribution with mean μ and standard deviation σ as the wage distribution. Under a shifting transformation (see Section 3.1). Let (μ, σ) be $(1, 1)$, the reserve wage be $\xi = 0.5$ and the number of observations be $N \in \{100, 400, 1600\}$. For every experiment, the number of repetition is 1000. Table 1 shows the finite sample performance of our modified likelihood estimator with the shifting transformation as well as a maximum estimator $\widehat{\xi} = \min_i w_i$. When we quadruple the sample size (from 100 to 400, or from 400 to 1600), the modified likelihood estimator reduces the standard deviation by half, which is consistent with the regular asymptotic properties of our estimator. On the other hand, $\widehat{\xi}$ shows evidence of superconsistency.

4.2. Auction Models

Consider the Power distribution $F(v) = v^{\theta}$ in first-price auctions, where $v \in [0, 1]$. Let the number of bidders be $I = 3$, and the parameter be

TABLE 1.

Monte Carlo Results: Search Models

	$N = 100$		$N = 400$		$N = 1600$	
	mean	std	mean	std	mean	std
$\widehat{\mu}$	1.0089	0.0900	1.0050	0.0457	1.0007	0.0220
$\widehat{\sigma}$	0.9907	0.0759	0.9956	0.0378	0.9988	0.0191
$\widehat{\xi}$	0.5185	0.0186	0.5048	0.0051	0.5013	0.0012

$\theta \in \{0.5, 1, 2\}$, and the number of observations be $N \in \{100, 400, 1600\}$. For every experiment, the number of repetition is $10,000$. Table 2 shows the finite sample performance of our modified likelihood estimator with the scaling transformation. When we quadruple the sample size (from 100 to 400, or from 400 to 1600), the modified likelihood estimator reduces the standard deviation by half, which is consistent with the regular asymptotic properties of our estimator.

TABLE 2.

Monte Carlo Results: Auction Models

θ		0.5	1	2
$N = 100$	mean	0.5103	1.0190	2.0377
	std	0.051	0.1034	0.2059
$N = 400$	mean	0.5024	1.0038	2.0107
	std	0.0255	0.0501	0.1008
$N = 1600$	mean	0.5006	1.0008	2.0022
	std	0.0125	0.0254	0.0499

4.3. Frontier Production Functions

Consider two inputs: capital and labor. Their logs are i.i.d. draws from a uniform distribution on $[0, 2]$. Let the true parameters be $\theta = (1, 0.5, 0.5)'$ and $\gamma = 1$. Thus, the production function has constant return to scale. Table 3 shows the finite sample performance of our estimator. When we quadruple the sample size (from 100 to 400, or from 400 to 1600), the modified likelihood estimator reduces the standard deviation by half, which is consistent with the regular asymptotic properties of our estimator.

TABLE 3.

Monte Carlo Results: Frontier Production Functions

	N=100		N=400		N=1600	
	mean	std	mean	std	mean	std
θ_0	1.0266	0.4143	1.0041	0.1983	0.9996	0.1009
θ_1	0.4884	0.3286	0.5011	0.1605	0.4965	0.0802
θ_2	0.4940	0.3194	0.4975	0.1600	0.5047	0.0814
γ	1.0368	0.1042	1.0105	0.0509	1.0025	0.0245

5. AN EMPIRICAL APPLICATION TO DOT PROCUREMENTS

In this section, we apply our method to data from Michigan Department of Transportation (MDOT) procurements. Up to now, we have assumed away covariates in auctions. In empirical applications, we may observe auction-specific covariates, such as engineer estimate, project type and duration et al., that may shift the distribution of bidder's private information. Therefore, we need to adapt our method to allow covariates. Guerre, Perrigne, and Vuong (2000) propose standard kernel smoothing over covariates. However, a fully nonparametric estimation approach may not be practical due to curse-of-dimensionality.

As an alternative, Haile, Hong, and Shum (2003) proposes a homogenization approach for incorporating covariates that became popular due to its convenience. See, e.g., Krasnokutskaya (2011), Bajari, Houghton, and Tadelis (2014), An (2017) and Liu and Luo (2017). In particular, they assume two competing models: an additive or multiplicative separable structure of the private information.[11] These structures are convenient because they are preserved by equilibrium bidding. However, the choice of structure has been mostly based on judgment calls rather than a formal model selection procedure. Note that this is a Vuong non-nested test of model selection. See Vuong (1989). As an illustration of the usefulness of our method, we consider a Vuong non-nested test of additive v.s. multiplicative separable observable heterogeneity in MDOT procurements.[12]

[11] Recently, Gimenes and Guerre (2022) propose a novel quantile regression approach, which is more flexible but does not suffer from curse-of-dimensionality.

[12] An existing alternative is the simulation-based selection method proposed in Li (2009). A formal comparison of the two methods is left for future research.

5.1. Data Sample

We collect data on the highway construction and maintenance procurements from MDOT during the first half of 2017.[13] For each auction, the data include the engineer's estimate, the number of bidders and the winning bid. We exclude auctions with a single bidder. Table 4 presents some summary statistics of the variables. The project size, measured by engineer estimate, ranges from about 40 thousands to over 60 millions. Therefore, it is important to incorporate covariates in our estimation.

We run a simple OLS regression of the logarithm of bid on the logarithm of the engineer estimate and the number of bidders. Consistent with the literature, we find that the fitting is very well with an R^2 of 98%. The coefficient on the logarithm of the engineer estimate shows that an 1% increase leads to 0.99% increase in the bid. On the other hand, the level of competition significantly reduces the bid.

TABLE 4.

Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Number of Bidders	382	4.47	2.52	2	17
Winning Bid	382	1.34E+06	3.55E+06	3.81E+04	5.99E+07
Engineer Estimate	382	1.48E+06	3.69E+06	3.94E+04	6.13E+07

5.2. Additive v.s. Multiplicative Separable Covariates

In this subsection, we introduce two estimators for the two competing parametric procurement auction models with additive and multiplicative separable auction-specific covariates, respectively. We maintain the assumption that the cost distribution belongs to the exponential family, which is commonly adopted for estimating a parametric auction model. See, e.g., Donald and Paarsch (1993) and Li (2010).

Multiplicative Separable Covariates

We first consider multiplicative separable covariates. Note that if x has an exponential distribution with rate $\gamma > 0$, $y = kx$ has an exponential distribution with rate γ/k, where $k > 0$. This motivates to consider the

[13]Procurement data from state deparment of transportation are widely used in the literature. See, e.g., Li and Zheng (2009), Krasnokutskaya (2011), Luo and Takahashi (2022), Gentry, Komarova, and Schiraldi (2023) and Kroft, Luo, Mogstad, and Setzler (2023).

exponential family of distributions for the cost in procurement auctions

$$f(c|z;\beta) = \gamma(z;\beta)\exp(-\gamma(z;\beta)c),$$

where $c > 0, \gamma(z;\beta) = \exp(\beta_0 + z'\beta_1) > 0$ and z represents auction-specific covariates. The winning bid is linear in the lowest cost $b(c) = c + \frac{1}{\gamma(z;\beta)(I-1)}$ and exponentially distributed

$$g_m(b;\beta) = \gamma(z;\beta)I \cdot \exp\Big(-\gamma(z;\beta)Ib + \frac{I}{I-1}\Big),$$

where $b > \frac{1}{\gamma(z;\beta)(I-1)}$. This support implies that $\log b + \log(I-1) > -\log\gamma(z;\beta) = -\beta_0 - z'\beta_1$.

We propose a two-step procedure to adapt our method. In the first step, we exploit the identification power of the support of the bid $[\frac{1}{\gamma(z;\beta)(I-1)}, +\infty)$. In particular, we propose a first-step estimator as a simple linear programming estimator for the parameters β:

$$\widetilde{\beta} = \arg \max_{\beta:\log b_i + \log I_i \geq -\beta_0 - z_i'\beta_1, \forall i} \beta_0,$$

See Smith (1994) for its properties.

In the second step, we propose a modified likelihood estimator based on the linear programming estimator $\widetilde{\beta}$. In particular, we consider a shifting transformation $y = x - \frac{1}{\gamma(z;\beta)(I-1)} + \underline{b}(z)$. The density of y becomes

$$\mathrm{g}_m(y;\beta) = g_m\Big(y + \frac{1}{\gamma(z;\beta)(I-1)} - \underline{b}(z);\beta\Big),$$

which leads to the induced loglikelihood function

$$\log \mathrm{g}_m(y;\beta) = \log(\gamma(z;\beta)I) - \gamma(z;\beta)I(y - \underline{b}(z)).$$

Thus, we can define our modified likelihood estimator:[14]

$$\widehat{\beta}_m = \arg\max_{\beta} \frac{1}{N}\sum_{i=1}^{N}\Big[(\beta_0 + z_i'\beta_1) - I_i \cdot \exp(\beta_0 + z_i'\beta_1)\cdot(b_i - \widehat{\underline{b}}_i)\Big],$$

[14]Absence of auction heterogeneity, the expected likelihood becomes $E[\log \mathrm{g}(b;\gamma)] - \text{const} = \log\gamma - \gamma I E[b + 1/(\gamma(I-1)) - \underline{b}]$, which leads to a simple MLE $\widehat{\gamma} = \frac{1/I}{\frac{1}{N}\sum_{i=1}^{N} b_i - \min_i b_i}$.

where $\widehat{\underline{b}}_i = \frac{1}{\exp(\widetilde{\beta}_0 + z_i'\widetilde{\beta}_0)\cdot(I_i-1)}$ is an estimate of the lower bound of the bid for auctions with covariates z_i based on the linear programming estimates.[15]

Additive Separable Covariates

An alternative specification of the model is to consider a benchmark model where

$$f(\epsilon) = \gamma \exp(-\gamma\epsilon),$$

and introduce additive separable covariates so that the cost satisfies

$$c = \beta_0 + z'\beta_1 + \epsilon.$$

Note that this additive separable structure is also preserved by equilibrium bidding. Moreover, the winning bid is exponentially distributed

$$g_a(b;(\beta,\gamma)) = \gamma I \exp\Big\{ -\gamma I\Big[b - (\beta_0 + z'\beta_1)\Big] + \frac{I}{I-1}\Big\},$$

where

$$b > \beta_0 + z'\beta_1 + \frac{1}{\gamma(I-1)}.$$

In sum, the unknown vector of parameters is $\theta \equiv (\beta,\gamma)'$.

Exploiting this lower boundary condition, we can also propose a first-step estimator as a simple linear programming estimator for the parameters:

$$(\widetilde{\beta},\widetilde{\gamma}) = \arg \max_{(\beta,\gamma):b_i \geq \beta_0 + \beta_1\cdot z_i + \frac{1}{\gamma(I_i-1)},\forall i} \beta_0,$$

In the second step, we can also propose a modified likelihood estimator based on the linear programming estimator $(\widetilde{\beta},\widetilde{\gamma})$. In particular, we consider a shifting transformation $y = x - k(\beta,\gamma) + \underline{b}(z)$, where $k(\beta,\gamma) = \beta_0 + z'\beta_1 + \frac{1}{\gamma(I-1)}$. The density of y becomes

$$\mathrm{g}_a(y;(\beta,\gamma)) = g(y + k(\beta,\gamma) - \underline{b}(z);(\beta,\gamma)),$$

which leads to the induced loglikelihood function

$$\log \mathrm{g}_a = \log(\gamma I) - \gamma I(b - \underline{b}(z)).$$

[15] The determination of this MLE can be simplified by successive maximization. In particular, for any given β_1, the optimal β_0 is $\log N - \log\Big[I_i \sum_{i=1}^N \exp(z_i'\beta_1)\cdot(b_i - \widehat{\underline{b}}_i)\Big]$.

Note that only γ is identified through the induced likelihood. Thus, we can define our modified likelihood estimator:

$$\widehat{\gamma}_a = \arg\max_{\gamma} \frac{1}{N}\sum_{i=1}^{N}\Big[\log\gamma - \gamma I_i\cdot(b_i - \widehat{\underline{b}}_i)\Big] = \frac{1}{\frac{1}{N}\sum_{i=1}^{N}[I_i\cdot(b_i-\widehat{\underline{b}}_i)]},$$

where $\widehat{\underline{b}}_i = \widetilde{\beta}_0 + z_i'\widetilde{\beta}_1 + \frac{1}{\widetilde{\gamma}(I_i-1)}$ is an estimate of the lower bound of the bid for auctions with covariates z_i based on the linear programming estimates.[16]

5.3. Estimation Results and Vuong Test

Table 5 presents our structural estimation results. We report both the linear programming estimates as well as the proposed maximum induced-likelihood estimator. Standard errors for the linear programming estimates are complicated to obtain, but the ones for the modified likelihood estimates are in parenthesis.

Under the multiplicative separable structure, note that the mean of the cost density function is $1/\gamma(z)$, and its logarithm equals $(-\beta_0-\beta_1 z)$. For interpretability of the estimated coefficients, we let $z = \log(\text{Engineer Estimate})$. Thus, θ_1 represents the percentage change of the expected cost corresponding to a 1% change of Engineer Estimate.

An important benefit of our estimator is that we can calculate the standard errors in a familiar way: $\widehat{\text{se}} = diag(\sqrt{\text{Hessian}^{-1}})$, where Hessian^{-1} is the inverse of the Hessian matrix, i.e. second derivatives of the objective function for the unconstrained likelihood maximization problem. Our estimation results show that a 1% increase in the Engineer Estimate leads to approximately 1% increase in the expected cost, as well. The standard error of $\widehat{\beta}_1$ indicates a strong significance.

We also estimate the model under the additive separable structure. The first-step and second-step estimates are reported in columns 3 (LP2) and

[16] Alternatively, we consider a scaling transformation $y = x\underline{b}(z)/k(\beta,\gamma)$, where $k(\beta,\gamma) = \beta_0 + z'\beta_1 + \frac{1}{\gamma(I-1)}$. The density of y becomes

$$\mathrm{g}(y;(\theta,\gamma)) = g(yk(\beta,\gamma)/\underline{b}(z);(\beta,\gamma))\cdot k(\beta,\gamma)/\underline{b}(z),$$

which leads to a modified likelihood estimator:

$$\widehat{\beta} = \arg\max_{\beta}\frac{1}{N}\sum_{i=1}^{N}\Big[\log\Big(\widetilde{\gamma}(\beta_0+\beta_1 z_i) + \frac{1}{(I_i-1)}\Big) - I_i(\frac{b_i}{\widehat{\underline{b}}_i} - 1)\cdot\widetilde{\gamma}(\beta_0 + z'\beta_1)\Big],$$

where $\widehat{\underline{b}}_i = \widetilde{\beta}_0 + z_i'\widetilde{\beta}_1 + \frac{1}{\widetilde{\gamma}(I_i-1)}$ is an estimate of the lower bound of the bid for auctions with covariates z_i based on the linear programming estimates. Note that only the coefficients for the covariates $\beta = (\beta_0,\beta_1)'$ are identified through this induced likelihood so we have replaced γ by the linear programming estimate $\widetilde{\gamma}$.

TABLE 5.

Estimated Structural Parameters

	LP1	ML1	LP2	ML2
β_0	1.7648	-1.0068 (0.7183)	-99,284	
β_1	-1.0934	-1.0088*** (0.0528)	0.5816	
γ			1.11E-05	4.40E-07*** (2.25E-08)

Note: *** $p < 0.01$

4 (ML2). For interpretability of the estimated coefficients, we let $z =$ Engineer Estimate. Note that the two competing structures have the same distributional assumption but different functional forms. Thus, they are strictly non-nested models. Following Vuong (1989), we can calculate the non-nested model selection test statistics. Define

$$\widehat{LR}_N = \sum_{i=1}^{N} \log \frac{g_m(b_i|z_i;\theta_m)}{g_a(b_i|z_i;\theta_a)} = -27.43,$$

where $\theta_m = \widehat{\beta}_m$ and $\theta_a = (\widetilde{\beta}_a, \widehat{\gamma}_a)'$.

First, the variance can be estimated as

$$\widehat{\omega}_N^2 = \frac{1}{N}[\sum_{i=1}^{N} \log \frac{g_m(b_i|z_i;\theta_m)}{g_a(b_i|z_i;\theta_a)}]^2 - [\frac{1}{N}\sum_{i=1}^{N} \log \frac{g_m(b_i|z_i;\theta_m)}{g_a(b_i|z_i;\theta_a)}]^2 = 1.69,$$

which is clearly not zero. Second, since the two models are strictly non-nested, we calculate the test statistic

$$\widehat{T}_N = N^{-1/2} \cdot \widehat{LR}_N / \widehat{\omega}_N = -1.07.$$

Note that $\Phi(\widehat{T}_N) = 14\%$, where $\Phi(\cdot)$ is the standard normal distribution. This suggests that both models explain the data equally well.

6. CONCLUSION

In boundary models, we propose a new estimator that maximizes an induced-likelihood function that is derived from the family chosen for the original likelihood function. In particular, we propose a simple transformation such that the transformed variable has the same support as the true distribution of the date generating process for any parameter value and has an expected likelihood function that maximizes at the true parameter value. As a result, our modified likelihood estimator has the regular asymptotic properties. We illustrate our method using three sets of models: search models, auction models and deterministic frontier production functions. Lastly, we apply our method to data from Michigan Department of Transportation procurements during the first half of 2017. After estimating the first-price auction model under additive and multiplicative separable observed heterogeneity, respectively, we implement a Vuong non-nested test of model selection. The test seems to suggest that both models explain the data equally well.

APPENDIX

A.1. MONTE CARLO RESULTS: INCORPORATING COVARIATES

We consider first-price procurement auctions where only winning bids are observed to the analyst and follow Li (2010) for the data generating process: the number of bidders is $I = 6$ and the auction-specific covariates z_ℓ, $\ell = 1, 2, \ldots, L$, are i.i.d. draws from a uniform distribution on $[0, 2]$. At the ℓ-th auction, the bidders draw their private costs from an exponential distribution with the density function

$$f(c|z_\ell) = \gamma(z_\ell;\theta)\exp(-\gamma(z_\ell;\theta)c),$$

where $c > 0, \gamma(z_\ell;\theta) = \exp(\beta_0 + \beta_1 z_\ell) > 0$. The true parameter value is $\theta_0 = (1, 0.5)'$.

Table 6 reports the results on the estimates of $\theta = (\beta_0, \beta_1)'$ using the three estimators: the linear programming estimator, the modified MLE and the indirect inference estimator (using OLS as the auxiliary model). For the indirect inference estimator, we follow closely Li (2010): (1) the first-stage OLS is regressing bid on a constant and auction-specific covariates z_ℓ; (2) the number of simulations is 1. Since the number of auxiliary parameters is equal to the number of the structural parameters, the indirect inference estimator is the solution to a nonlinear equation system.[17]

[17] We use the bid quantile function to simulate bids, i.e., $b(\alpha|z_\ell;\theta) = [\frac{I}{I-1} - \log(1-\alpha)]/[I\exp(\beta_0 + \beta_1 z_\ell)]$. For each sample, we redraw a sample of i.i.d. draws from a standard uniform distribution and use the same draws to calculate the simulated bids when the parameter varies. As a result, the simulated bids become smooth functions of

TABLE 6.

Comparing the Performance of Different Estimators

	$N = 100$		$N = 400$		$N = 1600$	
	mean	std	mean	std	mean	std
$\widetilde{\beta}_0^{LP}$	0.9839	0.0311	0.9966	0.0077	0.999	0.002
$\widetilde{\beta}_1^{LP}$	0.5006	0.029	0.4997	0.0072	0.5001	0.0018
$\widehat{\beta}_0^{mMLE}$	1.0261	0.211	1.0044	0.1046	1.0007	0.0505
$\widehat{\beta}_1^{mMLE}$	0.4978	0.1796	0.5025	0.0893	0.5016	0.0457
$\widehat{\beta}_0^{II}$	0.9961	0.1308	1.0033	0.0651	0.9987	0.0324
$\widehat{\beta}_1^{II}$	0.5053	0.1197	0.4978	0.0560	0.5011	0.0288

Since all three estimators are computationally efficient here, the computational advantage of the modified MLE is not significant. The indirect inference estimator seems to have some advantages in terms of bias and standard deviation in this example.[18] In particular, it reduces the standard deviation of the modified MLE by about 40% across the board. When we quadruple the sample size L (from 100 to 400, or from 400 to 1600), the modified likelihood estimator $\widehat{\theta}$ reduces the standard deviation by half, which is consistent with the regular asymptotic properties of our estimator. So as the indirect inference estimator. On the other hand, the standard deviation of the linear programming estimator $\widetilde{\theta}$ reduces to a quarter.

REFERENCES

Aigner, D. J., and S.-F. Chu, 1968. On estimating the industry production function. *The American Economic Review*, 826–839.

Amsler, C., M. Leonard, and P. Schmidt, 2013. Estimation and inference in parametric deterministic frontier models. *Journal of Productivity Analysis* **40(3)**, 293–305.

An, Y., 2017. Identification of first-price auctions with non-equilibrium beliefs: A measurement error approach. *Journal of Econometrics* **200(2)**, 326–343.

Aryal, G., M. F. Gabrielli, and Q. Vuong, 2021. Semiparametric estimation of first-price auction models. *Journal of Business and Economic Statistics* **39(2)**, 373–385.

Athey, S., and P. A. Haile, 2007. Nonparametric approaches to auctions. *Handbook of econometrics* **6**, 3847–3965.

the parameters because the bid quantile function is. This facilitates solving the nonlinear equation system.

[18]While the asymptotic distribution of the indirect inference estimator is complicated, the bootstrap can be used as an alternative in this example because each replication takes little time.

Bajari, P., S. Houghton, and S. Tadelis, 2014. Bidding for incomplete contracts: An empirical analysis of adaptation costs. *The American economic review* **104(4)**, 1288–1319.

Bronnenberg, B. J., S. K. Dhar, and J.-P. H. Dubé, 2011. Endogenous sunk costs and the geographic differences in the market structures of CPG categories. *Quantitative Marketing and Economics* **9(1)**, 1–23.

Chernozhukov, V., and H. Hong, 2004. Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* **72(5)**, 1445–1480.

Christensen, B. J., and N. M. Kiefer, 1991. The exact likelihood function for an empirical job search model. *Econometric Theory* **7(4)**, 464–486.

Donald, S. G., and H. J. Paarsch, 1993. Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *International Economic Review*, 121–148.

Flinn, C., and J. Heckman, 1982. New methods for analyzing structural models of labor force dynamics. *Journal of Econometrics* **18(1)**, 115–168.

Gentry, M., T. Komarova, and P. Schiraldi, 2023. Preferences and performance in simultaneous first-price auctions: A structural analysis. *The Review of Economic Studies* **90(2)**, 852–878.

Gimenes, N., and E. Guerre, 2022. Quantile regression methods for first-price auctions. *Journal of Econometrics* **226(2)**, 224–247.

Greene, W. H., 1980. On the estimation of a flexible frontier production model. *Journal of Econometrics* **13(1)**, 101–115.

Guerre, E., I. Perrigne, and Q. Vuong, 2000. Optimal Nonparametric Estimation of First-Price Auctions. *Econometrica* **68(3)**, 525–574.

Haile, P. A., H. Hong, and M. Shum, 2003. Nonparametric tests for common values at first-price sealed-bid auctions. Discussion paper, National Bureau of Economic Research.

Hill, J. B., and A. Shneyerov, 2013. Are there common values in first-price auctions? A tail-index nonparametric test. *Journal of econometrics* **174(2)**, 144–164.

Hirano, K., and J. R. Porter, 2003. Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support. *Econometrica* **71(5)**, 1307–1338.

Krasnokutskaya, E., 2011. Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies* **78(1)**, 293–327.

Kroft, K., Y. Luo, M. Mogstad, and B. Setzler, 2023. Imperfect competition and rents in labor and product markets: The case of the construction industry. Discussion paper, National Bureau of Economic Research.

Laffont, J.-J., and Q. Vuong, 1993. Structural econometric analysis of descending auctions. *European Economic Review* **37(2-3)**, 329–341.

Li, T., 2009. Simulation based selection of competing structural econometric models. *Journal of Econometrics* **148(2)**, 114–123.

Li, T., 2010. Indirect inference in structural econometric models. *Journal of Econometrics* **157(1)**, 120–128.

Li, T., and B. Zhang, 2015. Affiliation and entry in first-price auctions with heterogeneous bidders: an analysis of merger effects. *American Economic Journal: Microeconomics* **7(2)**, 188–214.

Li, T., and X. Zheng, 2009. Entry and competition effects in first-price auctions: theory and evidence from procurement auctions. *The Review of Economic Studies* **76(4)**, 1397–1429.

Liu, N., and Y. Luo, 2017. A Nonparametric Test for Comparing Valuation Distributions in First-Price Auctions. *International Economic Review* **58(3)**, 857–888.

Luo, Y., and H. Takahashi, 2022. Bidding for Contracts under Uncertain Demand: Skewed Bidding and Risk Sharing. Working Papers tecipa-732, University of Toronto, Department of Economics.

Murphy, K. M., and R. H. Topel, 2002. Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* **20(1)**, 88–97.

Paarsch, H. J., 1992. Deciding between the common and private value paradigms in empirical models of auctions. *Journal of econometrics* **51(1-2)**, 191–215.

Schmidt, P., 1976. On the statistical estimation of parametric frontier production functions. *The review of economics and statistics* **58(2)**, 238–39.

Smith, R. L., 1985. Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72(1)**, 67–90.

Smith, R. L., 1994. Nonregular regression. *Biometrika* **81(1)**, 173–183.

Sutton, J., 1991. Sunk costs and market structure: Price competition, advertising, and the evolution of concentration. MIT press.

Vuong, Q. H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 307–333.