Abstract

In this paper, we propose a new estimator for the distribution of valuations in first-price auctions with independent private valuations which applies in parametric and nonparametric settings. A distinguishing feature of our estimator is that its distribution is well-approximated by a sequence of normal distributions with consistently estimable covariance matrices. Consequently, our method provides simple inference for a wide array of objects of interest, which includes—but is not limited to—forming confidence intervals for the density of valuations at a fixed point, the optimal reserve price, the expected revenue of the auction and quantiles of the distribution of valuations as well as providing uniform confidence bands for the density of valuations. In first-price auctions, it is well-known that the support of the bid distribution depends on the parameters of the valuation distribution, which violates the standard regularity conditions for the asymptotic normality of the maximum likelihood estimator. We circumvent this issue by applying a simple modification to a method of moments estimator, which uses moments derived from the likelihood function. Simulations suggest our estimator and inference procedure perform well as our confidence bands provide empirical coverage close to nominal size and the mean squared error of functions of interest using our estimator compare favorably against alternatives in the literature. We demonstrate the usefulness of our approach in an application to timber auctions conducted by the United States Forest Service. To illustrate the flexibility of our inference procedure, we show how policy makers can use our results to form confidence sets for the welfare-maximizing reserve price when non-revenue considerations enter the welfare function.

---

*We are grateful for advice and encouragement from Joachim Freyberger, Jack Porter and Bruce Hansen. We would also like to thank Xiaoxia Shi, Mikkel Sølvsten, Ken Hendricks, Chris Sullivan, and Dan Quint and seminar participants at UW-Madison for helpful comments and discussions. Our research was supported by the computational resources provided by the Center for High Throughput Computing at UW-Madison.

†NERA Economic Consulting. Email: cristian.hernandez@nera.com.

‡Department of Economics, University of Wisconsin-Madison. Email: breeves@wisc.edu.

§The latest version can be found on the following website: sites.google.com/view/brandonreeves/research-1
1 Introduction

Economists are often interested in studying how changes in market design affect market outcomes. For auctions, this involves studying how changes in the rules of the auction, such as changing the reserve price policy, affect the results of the auction, such as the expected revenue generated by the auction or the probability of selling the object. Due to its widespread use as an allocation mechanism, such as auctions for mineral or resource rights, the rights to transmit signals on electromagnetic spectra, contracts for building or maintaining highways and selling online advertising space, there is a large literature on the econometrics of first-price auctions. While the literature on identification and consistent estimation of these models is well-developed, forming asymptotically valid confidence sets for general functions of the distribution of valuations in first price auctions has been a longstanding unresolved problem. Without a valid inference procedure, applied researchers are left without a formal method to assess the precision of their point estimates for functions of interest. A primary difficulty is that the object of interest is a feature of the distribution of unobserved valuations. As bids are the outcome of an unknown mapping applied to the valuations, standard approaches to characterize the limiting behavior of estimators do not apply even if one is willing to assume a parametric specification.

In this paper, we propose a new estimator for the distribution of valuations in first-price auctions which consists of a simple modification of a method of moments estimator with moments constructed from the log-likelihood function. The particular construction of the moments as well as the simple modification avoids the non-standard features of the first-price auction model. As a result, we can show the distribution of our estimator is well-approximated by a sequence of normal distributions. When combined with mild continuity conditions on the test-statistic, these distributional approximations allow us to produce \((1 - \alpha)100\%\) confidence sets for general functions of interest. This is the first result in the first-price auctions literature that allows researchers to form confidence sets for general functions. Although our results are not limited to such examples, our inference procedure allows the formation of confidence intervals for 1) the density of valuations, 2) quantiles of the valuations, 3) the expected revenue of the auction for a given reserve price and 4) the revenue-maximizing reserve price. As the our estimator is asymptotically normally distributed, it is computationally simple to form confidence sets using our estimator. For example, using the t-statistic, our 95\% confidence set is equal to the point-estimate plus/minus 1.96 times the estimated standard deviation.
Our estimator consists of a simple modification to a method of moments estimator where the estimating moments are given by asymptotically demeaned sample score functions. By using this criterion function, our estimator belongs to a general class of estimators which are constructed by applying a Newton-Rhapson step to a restricted minimizer of a sample criterion function. Intuitively, the application of the Newton-Rhapson step constructs our estimator as the unrestricted minimizer of a quadratic approximation of the criterion function around the restricted estimator.

Using similar arguments as those proposed in Ketz (2018) for parametric models, we show that under mild regularity conditions our proposed estimator is well-approximated by a sequence of normal distributions. Although we focus on the special case of a simple first-price auction model, our approach applies more generally to form asymptotically normally distributed estimators from parametric/nonparametric M-estimators where the criterion function may only be defined on a subset of the parameter space. Consequently, our results may apply to more elaborate first-price auction environments and may be of independent interest outside of the auction literature.

We examine the performance of our estimator and confidence bands in a Monte Carlo simulation using three criteria. First, we assess the finite sample coverage of our confidence bands for several commonly estimated functionals. Second, we compare the performance of our estimator to the method of moments (MM) and maximum likelihood (ML) estimators. This comparison gauges the cost, in terms of increased variance, of using our normally distributed estimator as opposed to the potentially non-normally distributed “restricted” alternatives. Third, we compare our estimator to estimators in the literature. In all three criteria, our estimator performs well. In the simulation, our confidence sets had coverage close to nominal size. Moreover, the variance of our estimator is only slightly larger than the variance of the MM and ML estimators. For example, for the optimal reserve price, our estimator had less than a 5% increase in the variance over the MM and ML estimators. Finally, our method compares favorably to alternatives in the literature. For example, the mean-squared error of our estimate of the optimal reserve price was 86.5% smaller than the estimator in GPV (2000) and 81.9% smaller than the estimator in Marmer and Shneyerov (2012).

To demonstrate the usefulness of our estimator and confidence bands, we include an application to sealed-bid, first-price auctions for timber rights conducted by the United States Forest Service. In these auctions, agents bid for the right to harvest timber from federally-managed lands. Using observational bid data, we nonparametrically estimate the density of valuations and present three inference results to show the flexibility of our approach. Specifically, we include 1) uniform confidence bands for the density of valuations, 2) a confidence interval for the revenue-maximizing
reserve price and 3) joint uniform confidence bands for the sale probability and the expected revenue of the auction both as functions of the reserve price. The first two results illustrate our approach can produce confidence bands for commonly estimated objects in the literature. The last feature shows how policymakers can use our approach to evaluate and make inferences about the potential trade-offs of a particular reserve price policy when non-revenue considerations are important. Moreover, if the welfare function of the agency were known, one could use our method to estimate and make inferences about the welfare optimizing reserve price. We find the revenue increase in moving from no-reserve to the revenue-optimal reserve price is small and results in a substantial reduction in the probability of a sale. The narrow confidence bands suggest policymakers may optimally set low reserve prices to mitigate the probability of not selling the harvesting rights. These considerations could be important if the Forest Service is concerned with responsibly managing potentially overgrown tracts or is attempting to increase the supply of timber to US sawmills, which are both stated objectives of the US Forest Service.

In the rest of this section, we explain how our paper fits into the literature. To keep the discussion concise, we only focus on papers which address the problem of inference in first-price auctions. As stated previously, the primary difficulty in the asymptotic analysis of estimators in first-price auction models is the unknown mapping between the observed bids and the primitives of the model, which are the valuations. This mapping complicates the formation of estimators and the derivation of their asymptotic distributions as it often violates standard arguments used to approximate the limiting behavior of the estimator. As a result, there is no existing method which can form confidence sets for general functions of interest in nonparametric settings.

In an early paper in the literature, Donald and Paarsch (1993) demonstrated the first-price auction model violates the regularity conditions for the maximum likelihood estimator as the support of the bid distribution depends on the parameters of the valuation distribution. Specifically, for first-price auctions with interval-supported valuations, the support of the bid distribution is an interval where the rightmost endpoint of the support is a function of the parameters of the valuation distribution. As a result of this non-standard feature, one cannot use standard asymptotic arguments to establish the limiting distribution of the parametric maximum likelihood estimator.

Early efforts in the literature focused on alternatives to the parametric maximum likelihood estimator. One alternative is the piecewise pseudo maximum likelihood estimator of Donald and Paarsch (1993). Their method assumes one can partition the parameter $\theta$ into a scalar, $\theta_1$, and the remaining parameters, $\theta_2$, where the rightmost support point of the bid distribution is strictly
increasing in $\theta_1$. For any value of $\theta_2$, they set $\theta_1$ to constrain the rightmost point of the bid distribution to equal the largest observed bid. Then, they obtain their estimator by maximizing the likelihood over $\theta_2$ after concentrating out $\theta_1$. Under mild regularity conditions, their estimator for $\theta_2$ is asymptotically normal. Laffont et al. (1995) proposed a second alternative based on simulated method of moments which estimates $\theta$ by minimizing a nonlinear least squares objective function measuring the distance between sample moments and simulated moments using $\theta$. By suitably adjusting the criterion function, their estimator is consistent using a fixed number of simulations as the sample size tends to infinity. Under some regularity conditions, which includes the potentially strong assumption that the moments identify the true parameter, they show their estimator is asymptotic normal. In contrast to our result, these papers rely on parametric assumptions and it is not immediately clear whether these arguments can be extended to nonparametric settings.

In GPV (2000), the authors propose a nonparametric, two-step kernel estimator for the distribution of valuations. The key insight of their estimator is that if the density of bids were known, one can recover the valuation which generated a given bid. To implement this observation, their estimator uses a first-step kernel estimate of the bid-distribution to construct plug-in estimates of the valuations, which they call pseudo-valuations. In the second step, they use the pseudo-valuations to produce their kernel-estimate of the density of valuations. A limitation of this approach is that it requires the user to specify number of bandwidth choices when the observed auctions involve distinct numbers of participants. Until recently, the asymptotic distribution of the GPV (2000) estimator was unknown. Ma et al. (2018) show the GPV (2000) estimator for the density of valuations at an interior point is asymptotically normal. While this resolved a longstanding question, the practical importance of confidence bands for the density of valuations is not clear. Often, the density of valuations is only of secondary interest, and it is unclear if one can extend this result to general functions of interest which may depend upon the entire distribution of valuations.

Marmer and Shneyerov (2012) propose a nonparametric estimator for the density of valuations based on the observation that the monotonicity of the bidding strategy implies quantiles of the valuations are known functions of the quantiles of the bids. To construct their estimator, they estimate quantiles of bids and use this mapping between the quantiles to construct quantiles of the valuation distribution. By inverting the quantile function and differentiating, they obtain their estimator for the valuation density. As it avoids the construction of pseudo-values, their estimator only involves one kernel estimator. This allows them to show their estimate for the density of valuations at an interior point is asymptotically normally distributed. While they do not establish the asymptotic
distribution of the plug-in estimator, they provide a method for constructing confidence sets for the optimal reserve price by inverting a test. As a result, these confidence sets can be arbitrary subsets of the valuation space and are generally not intervals. Furthermore, it is unclear if it is possible to use their results to form confidence sets for general functions of interest.

Our paper is also related to the literature on inference under shape-restrictions. We briefly discuss the connection to three related papers. In a parametric model, Andrews (1999) derives the asymptotic distribution for restricted estimators under binding shape-restrictions. When a restriction binds, he shows the asymptotic distribution of the restricted estimator is the projection of a normal distribution on to the restricted parameter space. Ketz (2018) studies a parametric model similar to Andrews (1999), but to avoid the non-normality of the restricted estimator, he proposes an “unrestricted” estimator obtained by applying a Newton-Rhapson step to the restricted estimator. Among other things, he shows this estimator is asymptotically normally distributed. We use a similar methodology as in Ketz (2018) to establish the approximate normality for the class of nonparametric problems we consider. Finally, Freyberger and Reeves (2018) provide a uniformly valid inference procedure for a growing parameter vector subject to binding or drifting-to-binding constraints. While we use similar technical arguments, the objective of the current paper differs considerably from Freyberger and Reeves (2018). Whereas their focus is on using the non-normally distributed restricted estimators as a basis for inference, our approach is concerned with constructing normally distributed unrestricted estimators from the restricted estimators. Although the confidence sets formed using the restricted estimator are generally smaller than those based on the unrestricted estimator, obtaining the restricted confidence bands can be computationally challenging. In contrast, the asymptotic normality of our unrestricted estimator allows the formation of standard confidence bands which are simple to obtain and easy to report.

The rest of the paper is organized as follows. Section 2 describes the auction environment and a description of its non-standard features. Section 3 introduces our estimator and high-level conditions under which our tests control size. Section 4 contains low-level sufficient conditions for a simple auction model. Section 5 contains the simulation study. Section 6 contains the empirical application. Section 7 concludes and discusses avenues for extensions. Additionally, we include an appendix of figures/tables in section 8 and a technical appendix in section 9. For the remainder of the paper, denote the density of bids and valuations by $f$ and $g$, respectively. For a positive definite and symmetric matrix $A$ and vector $x$, let $\|x\|_A^2 = x'Ax$. Moreover, let $\|A\|$ denote the Frobenius norm of the matrix $A$ and $\|A\|_S$ denote the spectral norm of the matrix $A$. 
2 The Auction Environment

To introduce the non-standard features of the model and show how our proposed estimator avoids these difficulties, our discussion focuses on a simple first-price auction model. In subsection 2.1 we outline the simple auction model we consider, and in subsection 2.2 we discuss the non-standard features. To simplify the presentation of the arguments for this section, we temporarily assume a parametric specification in which \( f_0(v) = f(v, \theta_0) \) for some \( \theta_0 \in \Theta \). In section 3.3 we return to the general environment which allows for both parametric and nonparametric specifications.

2.1 Description of the Auction Model

In the model we consider, a seller is auctioning a single object through a first-price auction in which valuations are private information and bids are submitted simultaneously. In the auction, the object is awarded to highest bidder and the winner pays their bid. We assume there is a non-binding reserve price so that no winning bid is rejected by the seller. Agents draw independent valuations from a distribution function with a density given by \( f(\cdot, \theta_0) \) with known support on the compact interval \([v, \overline{v}]\). When submitting bids, agents only know their valuation, the number of players against which they are bidding and the distribution \( f(\cdot, \theta_0) \). As they do not know the valuations of the other participants, agents bid to maximize expected utility conditional on their own valuation. We assume risk-neutrality so agents bid to maximize expected conditional payout.

The bidding function, denoted \( b(v, \theta, p) \), specifies the bid which maximizes the expected payout for an agent with valuation \( v \) when bidding in an auction with \( p \) participants (including the agent) when valuations are drawn from the density \( f(\cdot, \theta) \). The bidding strategy is given by

\[
b(v, \theta, p) = v - \int_v^\infty \frac{F(t, \theta)^{p-1} dt}{F(v, \theta)^{p-1}}.
\]

Under the assumption that \( f(v, \theta) > 0 \) for all \( v \in [v, \overline{v}] \) the bidding function is strictly increasing in valuations and thus admits an inverse bid function, denoted as \( \eta(b, \theta, p) \). As the inverse bid function must also be monotonic, we can derive the distribution of bids, \( G(b, \theta, p) \), as

\[
G(b, \theta, p) = P(b(v, \theta, p) \leq b) = P(v \leq \eta(b, \theta, p)) = F(\eta(b, \theta, p), \theta).
\]

Furthermore, if \( f(\eta(b, \theta, p), \theta) > 0 \), then \( \eta(b, \theta, p) \) is differentiable for all \( b \) in the interior of the range-space of the bidding function. Therefore, we can obtain the density of bids, denoted \( g(b, \theta, p) \) by differentiating equation (2) with respect to \( b \). By the chain-rule and the implicit function
theorem, for any $b$ in the range-space of $b(\cdot, \theta, p)$ we have

$$g(b, \theta, p) = \frac{\partial G(b, \theta, p)}{\partial b} = f(\eta(b, \theta, p)) \frac{\partial \eta(b, \theta, p)}{\partial b} = \frac{F(\eta(b, \theta, p), \theta)}{(p-1)(\eta(b, \theta, p) - b)}$$

where the last equality uses the implicit function theorem. Lastly, the density of bids is zero for any $b$ not in the range-space of the bidding function. In order to simplify the notation and presentation of the arguments, we assume we observe all bids from $L$ auctions with a fixed number of participants, $p$. As a result, we observe a sample of $n = pL$ total bids. It is easy to relax these assumptions to accommodate situations with a of heterogeneous number of participants and cases where we only observe the winning bid (i.e. the transaction price).

### 2.2 Illustration of the Non-Standard Features

Given expression (3), an intuitive estimator of $\theta$ is the maximum (log-)likelihood estimator. Under standard regularity conditions, maximum likelihood estimators are asymptotically normally distributed and asymptotically efficient. Unfortunately, it is possible to demonstrate these regularity conditions do not hold for the first-price auction model. In this section, we briefly discuss these non-standard features which motivates the particular construction of our estimator in section 3.

As illustrated in Donald and Paarsch (1993), the strategic behavior of auction participants induces dependence between the support of the bid distribution and the true parameter even if the support of valuation distribution is independent of the true parameter. Heuristically, the maximum bid rationalized by the auction model is submitted by the player with valuation $\bar{v}$, and this agent strategically responds to the entire shape of the valuation distribution. To see this, define

$$\bar{b}(\theta, p) = \lim_{v \to \bar{v}} b(v, \theta, p) = \bar{v} - \int_{\bar{v}}^{\infty} F(t, \theta)^{p-1} dt,$$

which denotes the largest bid the model can rationalize in auctions with $p$ participants when the valuation distribution is given by $f(\cdot, \theta)$. The second term in the expression reflects the utility-maximizing considerations of the agent with the largest valuation as they attempt to capture the largest (expected) surplus of winning the auction. As a result of these strategic considerations, the support of the bid distribution depends non-trivially on the shape of the valuation distribution. This dependence—which is the primary source of the non-standard features of the model—is not specific to the simple auction model we consider and is a feature commonly encountered in more general auction environments such as in affiliated and common value auction models.

One of the primary complications introduced by the parameter-dependent support is that estimators based on the log-likelihood function behave like shape-restricted estimators with potentially
non-standard distributions. To see this, fix a sample of bids \( \{\{b_i\}_{i=1}^p\}_{i=1}^L \) and consider evaluating the log-likelihood for an arbitrary value of \( \theta \). By equation (4), we see the largest bid the model can rationalize given \( \theta \) is \( \bar{b}(\theta, p) \). If any bid in the sample exceeds \( \bar{b}(\theta, p) \), the parameter \( \theta \) could not have generated the given sample as no agent would rationally bid above \( \bar{b}(\theta, p) \) for any valuation in \([v, \bar{v}]\) if valuations were distributed as \( f(\cdot, \theta) \). Therefore, the sample likelihood at \( \theta \) is zero, which implies the sample log-likelihood is not well-defined for this \( \theta \). If we denote \( \hat{\Theta}_R \equiv \{ \theta \in \Theta \Bigr| \frac{\partial}{\partial \theta} b(\theta, p) \neq 0 \} \), the log-likelihood function is only well-defined on \( \hat{\Theta}_R \). Furthermore, as the sample size grows, the maximum observed bid converges almost surely to \( \bar{b}(\theta_0, p) \). Therefore, the log-likelihood is not well-defined for any parameter \( \theta \) such that \( \bar{b}(\theta, p) < \bar{b}(\theta_0, p) \) with probability approaching one for almost all sample paths.\(^1\) Consequently, if \( \bar{b}(\theta, p) \) is continuously differentiable in \( \theta \) with \( \frac{\partial}{\partial \theta} b(\theta, p) \neq 0 \), any open-ball of \( \theta_0 \) contains points \( \theta \) for which \( \bar{b}(\theta, p) < \bar{b}(\theta_0, p) \).\(^2\) As a result, standard asymptotic arguments which assume \( \theta_0 \) is an interior point in the parameter space may provide poor approximations to finite sample distribution of the estimator.

A second complication introduced by the parameter-dependent support is that one can no longer freely interchange the order of integration and differentiation. Freely interchanging the order of these operations is used to establish important properties of the maximum likelihood estimator. In particular, this property is used to show the zero expected value of the score function, which is crucial to establishing the asymptotic normality of the estimator. By applying the Leibniz integral rule we can obtain the closed-form expression for the expected value of the score function

\[
E_{\theta_0}\left( \frac{\partial \log(g(b, \theta_0, p))}{\partial \theta} \right) = -g(\bar{b}(\theta_0), \theta_0, p) \frac{\partial \bar{b}(\theta_0, p)}{\partial \theta} = \frac{\partial \bar{b}(\theta_0, p)}{\partial \theta} \int_{v}^{\bar{v}} F(t, \theta_0)^{p-1} dt \equiv \mu(\theta_0, p).
\]

Further, the inability to freely interchange the order of integration and differentiation invalidates the standard proof establishing the positive-definiteness of the Hessian matrix of the log-likelihood.

\(^1\)For such a \( \theta \), we may find an \( \eta > 0 \) with \( \bar{b}(\theta, p) < \bar{b}(\theta_0, p) - \eta \). The log-likelihood is not defined if there is a bid in the interval \([\bar{b}(\theta_0, p) - \eta, \bar{b}(\theta_0, p)]\). As \( g(b, \theta_0, p) > 0 \) for all \( b \in [v, \bar{b}(\theta_0, p)] \) with probability approaching one we observe at least one bid in this interval conditional on almost all sample paths.

\(^2\)\( F(v, \theta) \) continuously differentiable in \( \theta \) almost everywhere implie \( \bar{b}(\theta, p) \) is continuously differentiable for any \( p \).
3 Our Proposed Estimator and Testing Procedure

3.1 Description of Our Estimator

To address the non-standard features of the log-likelihood in section 2, our estimator consists of a simple modification of a method of moments estimator. The moments we use correspond to re-centered sample score functions. Specifically, our criterion function is

\[ Q_n(\theta) = m_n(\theta) m_n(\theta) \]

where

\[ m_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\partial \log(g(b_i, \theta, p))}{\partial \theta} - \mu(\theta, p) \right) \]

and \( \mu(\theta, p) \) is given in equation (6) so that \( E_{\theta}(m_n(\theta)) = 0 \). Due to the restricted-domain of the score, \( m_n(\theta) \) and thus \( Q_n(\theta) \) are not defined outside of \( \hat{\Theta}_R \). As a result, the minimizers of \( Q_n(\theta) \) will be shape-restricted estimators with non-normal asymptotic distributions. To prevent the effects of the boundary of \( \hat{\Theta}_R \) from entering the asymptotic distribution, we define our estimator to be

\[ \hat{\theta}_n = \hat{\theta}_{mm} + \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta \partial \theta'} \right)^{-1} \left( \frac{\partial Q_n(\hat{\theta}_{mm})}{\partial \theta} \right) \]

where \( \hat{\theta}_{mm} = \arg \min_{\theta \in \hat{\Theta}_R} n Q_n(\theta) \), (7)

which corresponds to a Newton-Rhapson step applied to \( \hat{\theta}_{mm} \). The construction of the estimator only requires expressions for the density and support of the bid distribution as functions of \( \theta \), which are available in many auction models. In the application in section 6 we show

To see how this modification avoids complications associated with the boundary of \( \hat{\Theta}_R \), note that \( \hat{\theta}_n \) minimizes a quadratic approximation to \( Q_n(\theta) \) centered at \( \hat{\theta}_{mm} \) (see equation (13) for details). By minimizing the quadratic approximation to \( Q_n(\theta) \) rather than minimizing \( Q_n(\theta) \), our estimator can take values outside of the restricted set. Consequently, the boundary effects are not present in the asymptotic distribution of our estimator. As we see in section 3.2, the use of re-centered moments in \( Q_n(\theta) \) and the Newton-Rhapson step address the two non-standard features of the model outlined in section 2.2 and results in the asymptotic normality of the estimator.

Although we introduced our estimator using the simple first-price auction model, the construction of “unrestricted” estimators from restricted estimators as in equation (7) can apply in settings with general criterion functions \( Q_n(\theta) \). This construction is useful when we seek normally distributed estimators in structural econometric models which restrict the set of parameters consistent with the data. In addition to the simple auction environment, similar restrictions appear in other first-price auction models such as affiliated and common value models, more elaborate independent private value models (e.g. with binding reserve prices, risk aversion, etc.) and in models outside the auction literature. A set of sufficient conditions for this method to produce asymptotically normal estimators is given in section 3.3.
3.2 Heuristic Outline of Asymptotic Normality

In this subsection, we provide a heuristic discussion of the arguments establishing the asymptotic normality of our proposed estimator. The arguments in this section are informal and are only intended to motivate the formal assumptions appearing in section 3.3. To illustrate the argument in a simple environment, we assume a parametric model in which $f_0(v) = f(v, \theta_0)$ for some $\theta_0 \in \Theta$.

Recall, from the previous section we have defined

$$
\hat{\theta}_{mm} = \arg \min_{\theta \in \hat{\Theta}} n Q_n(\theta).
$$

Due to the presence of the restriction, we cannot analyze the distribution of $\hat{\theta}_{mm}$ using the standard approach of linearizing first-order conditions. Instead, we follow the standard approach in the literature on inference under binding shape restrictions, see for instance Andrews (1999). These approaches start with a second-order Taylor expansion of $Q_n(\theta)$ about $\theta_0$. Assuming the Taylor-expansion holds exactly (c.f. Assumption 1 in section 3.3), we have

$$
n Q_n(\theta) = n Q_n(\theta_0) + \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta'} \sqrt{n} (\theta - \theta_0) + \frac{1}{2} \sqrt{n} (\theta - \theta_0)' \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} \sqrt{n} (\theta - \theta_0).
$$

Re-arranging this expression gives

$$
n Q_n(\theta) = n Q_n(\theta_0) - \frac{1}{2} D_n' J_n^{-1} D_n + \frac{1}{2} \| \sqrt{n} (\theta - \theta_0) - J_n^{-1} D_n \|^2_{J_n}
$$

where

$$
D_n = \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \quad \text{and} \quad J_n = \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'}.
$$

Substituting (9) into (8) and recognizing the first two terms in (9) do not depend on $\theta$ gives

$$
\hat{\theta}_{mm} = \arg \min_{\theta \in \hat{\Theta}} \| \sqrt{n} (\theta - \theta_0) - J_n^{-1} D_n \|^2_{J_n}.
$$

By using re-centered moments with $\mathbb{E}_{\theta_0}(m_n(\theta_0)) = 0$, $\sqrt{n} m_n(\theta_0)$ is $\frac{1}{\sqrt{n}}$ times a sum of mean-zero random variables, so $\sqrt{n} m_n(\theta_0)$ is asymptotically normally distributed by the central limit theorem under mild moment conditions. In conjunction with the law of large numbers, this shows

$$
D_n = \frac{\partial m_n(\theta_0)}{\partial \theta} \sqrt{n} m_n(\theta_0) \overset{d}{\to} N(0, V_D) \quad \text{and} \quad J_n \overset{p}{\to} J
$$

for symmetric matrices $V_D$ and $J$ (c.f. Assumption 2 in section 3.3). By performing the change of variables $\lambda = \sqrt{n} (\theta - \theta_0)$ in equation (10) and assuming non-singularity of $J$ (c.f. Assumption 6 in
section 3.3, we can see that $\sqrt{n} (\hat{\theta}_{mm} - \theta_0)$ is approximately distributed as the projection of the normal random variable $N(0, J^{-1}V_DJ^{-1})$ on to the space, $\sqrt{n} (\hat{\Theta}_R - \theta_0)$. The use of re-centered moments is crucial for establishing the normality in (11) (for further discussion, see section 3.4). Although one could use this result for hypothesis testing, the distribution of the estimator is non-standard, which may make forming confidence sets computationally difficult. As we seek a normally distributed estimator for which standard inference procedures apply, we take our estimator to be a simple modification of the estimator in equation (10).

To introduce the simple modification and show how it restores normality, note that the source of the non-standard features of the distribution of $\hat{\theta}_{mm}$ is the presence of the restriction in equation (10). This restriction is due to the limited domain on which the log-likelihood, and thus our moment, is defined. While $Q_n(\theta)$ is only defined for a subset of the parameter space, the right-hand side of equation (10) is a quadratic function of $\theta$ which is defined for all of $\Theta$. Minimizing this quadratic function over $\Theta$, yields the estimator $\theta_n = \theta_0 + J_n^{-1} \frac{1}{\sqrt{n}} D_n$, which corresponds to a Newton-Rhapson step applied to $\theta_0$. Re-arranging this and using the result in equation (11) gives

$$\sqrt{n} (\theta_n - \theta_0) = J_n^{-1}D_n \xrightarrow{d} N(0, J^{-1}V_DJ^{-1}).$$

Therefore, the Newton-Rhapson step effectively inverts the projection operator in equation (10) and recovers the component of the restricted estimator which is asymptotically normal.

Although it results in an asymptotically normally distributed estimator, the above method is infeasible as it corresponds to minimizing the quadratic expansion of $Q_n(\theta)$ centered at $\theta_0$ which is unknown. If we minimize the quadratic expansion of $Q_n(\theta)$ which is centered at $\hat{\theta}_{mm}$ we obtain our estimator which is the plug-in version of the infeasible estimator. Specifically, we get

$$\hat{\theta}_n = \hat{\theta}_{mm} + J_n^{-1} \frac{1}{\sqrt{n}}D_n = \hat{\theta}_{mm} + \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta \partial \theta'} \right)^{-1} \left( \frac{\partial Q_n(\hat{\theta}_{mm})}{\partial \theta} \right).$$

Although it is not immediately clear from equation (13), it is possible to show that under mild continuity conditions on the second derivative of $Q_n(\theta)$ (c.f. Assumption 4 in section 3.3) the feasible and infeasible estimators share the same asymptotic distribution. Specifically, we have

$$\sqrt{n} (\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, J^{-1}V_DJ^{-1}).$$

Intuitively, the continuity of the second derivative implies that local perturbations in the centering point for the quadratic expansion have a negligible impact on the quadratic expansion. As a result, centering the expansion around $\hat{\theta}_{mm}$ versus $\theta_0$ is asymptotically irrelevant. In the next section, we provide conditions which formalize this argument.
3.3 General Theory

In this section, we describe the testing procedure and include a set of high-level conditions under which the test asymptotically controls size. To allow for potential applications outside of the simple auction model, we present assumptions for a general criterion function $Q_n$ and test-statistic $T$. The cost of this generality, however, is the high-level nature of these assumptions. To accommodate practitioners, we provide a set of low-level sufficient conditions for the simple auction environment in the next section as an illustration of how to verify assumptions in this section.

Let $\Theta \subseteq \mathbb{R}^{K_n}$ denote the parameter space. For parametric models, $K_n$ is a fixed integer with $f_0(v) = f(v, \theta_0)$ for some $\theta_0 \in \Theta$. For nonparametric models, we take $\theta_0 \in \mathbb{R}^{K_n}$ such that $f_0(v) \approx f(v, \theta_0)$ with $K_n$ increasing in the sample-size. For a function(al) $\psi(\cdot)$, we consider tests of the null-hypothesis $H_0 : \psi(\theta_0) = \psi$. This may be a general functional such as the optimal reserve price, the expected revenue of the auction, quantiles of the distribution, etc.

To describe the test, let $\hat{\Sigma}_n$ be an estimate of the asymptotic variance of $\sqrt{n} (\hat{\theta}_n - \theta_0)$ and denote the test statistic by $T(\sqrt{n} (\hat{\theta}_n - \theta_0), \hat{\Sigma}_n)$. Some typical choices of the test statistic would be the t-statistic or the Wald test statistic. Let $c_{\alpha,n}(\hat{\Sigma}_n)$ denote the $1 - \alpha$ quantile of the distribution of $T(Z_n, \hat{\Sigma}_n)$ where $Z_n \sim N(0, \hat{\Sigma}_n)$. For example, using the t-test $c_{\alpha,n}(\hat{\Sigma}_n) = Z_{1 - \frac{\alpha}{2}}$ and using the Wald test gives $c_{\alpha,n}(\hat{\Sigma}_n) = \chi_{1 - \alpha, K_n}$. The test rejects $H_0$ if and only if $T(\sqrt{n} (\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) > c_{\alpha,n}(\hat{\Sigma}_n)$.

To form confidence sets, we invert the test by collecting the values of the null-hypothesis which are not rejected. In contrast to the standard case for shape-restricted estimators, our estimator is asymptotically normally distributed which implies the critical value is independent of $\theta_0$. As a result, the test inversion is algebraically simple and results in the standard confidence bands. For example, using the t-statistic yields the standard 95% confidence interval for $\psi(\theta_0)$ given by

$\left[ \psi(\hat{\theta}_n) - 1.96 \frac{\hat{\sigma}_\psi}{\sqrt{n}}, \psi(\hat{\theta}_n) + 1.96 \frac{\hat{\sigma}_\psi}{\sqrt{n}} \right]$ \text{ where } \hat{\sigma}_\psi^2 = \frac{\partial \psi(\hat{\theta}_n)}{\partial \theta \prime} \hat{\Sigma}_n \frac{\partial \psi(\hat{\theta}_n)}{\partial \theta}.

The following assumptions are sufficient to guarantee that tests which reject if and only if $T(\sqrt{n} (\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) > c_{\alpha,n}(\hat{\Sigma}_n)$ asymptotically control size. After stating the assumptions and the main theorem, we discuss the content and importance of each of the high-level assumptions. For the following assumptions, let $\epsilon_n$ be a sequence of positive numbers converging to zero where we discuss the role of $\epsilon_n$ after stating the assumptions and main result. Let $\mathcal{P}$ denote the set of probability models over which the following assumptions hold.

---

3We have suppressed the dependence of $\theta_0$ on $n$ for convenience.
Assumption 1. Quadratic Approximation There exists symmetric matrix $J_n$ and a random variable $\theta_n$ satisfying the following assumptions such that

$$\sqrt{n} (\hat{\theta}_{mm} - \theta_0) = \arg\min_{\lambda \in \sqrt{n} (\hat{\theta}_R - \theta_0)} \| \lambda - \sqrt{n} (\theta_n - \theta_0) \|_{J_n} + r_n$$

where $r_n = o_p(\varepsilon_n)$ uniformly over $\mathcal{P}$.

Assumption 2. Normality Approximation There exists symmetric matrices $\Sigma_n$ and a sequence of random variables $Z_n \sim N(0, \Sigma_n)$ such that $\|\sqrt{n} (\theta_n - \theta_0) - Z_n\| = o_p(\varepsilon_n)$ uniformly over $\mathcal{P}$.

Assumption 3. Covariance Estimation $\|\hat{\Sigma}_n - \Sigma_n\| = o_p(\varepsilon_n K_n^{-\frac{1}{2}})$ uniformly over $\mathcal{P}$.

Assumption 4. Continuity of the Criterion Function For any $\theta_1$ and $\theta_2$, we have

$$\left\| \frac{\partial Q_n(\theta_1)}{\partial \theta \partial \theta'} - \frac{\partial Q_n(\theta_2)}{\partial \theta \partial \theta'} \right\| \leq L_n \|\theta_1 - \theta_2\|$$

where $L_n = o_p(\frac{\varepsilon_n \sqrt{n}}{K_n})$ uniformly over $\mathcal{P}$.

Assumption 5. Continuity of the Test Statistic For all symmetric positive definite matrices $\Sigma_1$ and $\Sigma_2$ (with $\Sigma_1$ and $\Sigma_2$ comparable) with all eigenvalues in $[\frac{1}{B}, B]$ for some $B < \infty$ there exists a constant $C$, possibly depending on $B$, such that uniformly over $\mathcal{P}$

$$|T(z_1, \Sigma_1) - T(z_2, \Sigma_1)| \leq C\|z_1 - z_2\| \quad \text{and} \quad |T(z_1, \Sigma_1) - T(z_1, \Sigma_2)| \leq C\|z_1\|\|\Sigma_1 - \Sigma_2\|.$$

Assumption 6. Bounded Eigenvalues Uniformly over $\mathcal{P}$ the eigenvalues $\Sigma_n$ and $J_n$ are bounded and bounded away from zero.

Assumption 7. Anti-Concentration There exists a $\delta \in (0, \alpha)$ such that for all $\beta \in [\alpha - \delta, \alpha + \delta]$,

$$\sup_{P \in \mathcal{P}} |P(T(Z_n, \Sigma_n) \leq c_{\beta,n}(\Sigma_n) - \varepsilon_n) - (1 - \beta)| \rightarrow 0$$

and

$$\sup_{P \in \mathcal{P}} |P(T(Z_n, \Sigma_n) \leq c_{\beta,n}(\Sigma_n) + \varepsilon_n) - (1 - \beta)| \rightarrow 0.$$

Theorem 1. Suppose Assumptions 1-7 hold. Then

$$\sup_{P \in \mathcal{P}} |P\left(T\left(\sqrt{n} (\hat{\theta}_n - \theta_0), \hat{\Sigma}\right) \leq c_{\alpha,n}(\hat{\Sigma})\right) - (1 - \alpha)| \rightarrow 0.$$
Remark 1: This theorem states that the indicated testing procedure asymptotically controls size uniformly over $\mathcal{P}$. This immediately implies that the resulting confidence bands also control size. The proof of theorem (1) is contained in the appendix. The proof consists of two parts. In the first part, we establish that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically close to the random variable $Z_n$ described in assumption (2). In the second part of the theorem, we show that the impact of the two approximation errors (i.e. the approximation error introduced in Assumptions (2) and (3)) result in an asymptotically negligible distortion of the rejection probability of the test.

Remark 2: Several comments are in order about the assumptions. Assumptions (1) and (2) ensure the MM estimator behaves approximately as the projection of a normal random variable on to the sample restriction space. In conjunction with assumption (4), these assumptions establish the feasible plug-in estimator $\hat{\theta}_n$ and the infeasible estimator $\theta_n$ described in section 3.2 share the same asymptotic distribution. Assumption (5) is a mild continuity condition on the test-statistic which when combined with assumption (6) ensures the approximation errors in assumptions (2) and (3) have an asymptotically negligible impact on the distribution of the test-statistic. Assumption (6) can be relaxed to allow for growing/shrinking eigenvalues at polynomial rates at the cost of slightly stronger assumptions on the rates at which the other approximation errors converge to zero. In light of the first six assumptions, we can show the $1 - \alpha$ quantile of the distribution of $T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n)$ is at most $\varepsilon_n$ distance away from the corresponding quantile of $T(Z_n, \Sigma_n)$. As the distribution of $T(Z_n, \Sigma_n)$ changes with $n$, the distribution of the test-statistic may become steeper as $n$ increases. Consequently, we must ensure that $\varepsilon_n$ shrinks faster than the rate at which the distribution of $T(Z_n, \Sigma_n)$ is becoming steep at the $1 - \alpha$ quantile, which is precisely the role of Assumption (7). This type of assumption is commonly referred to as an anti-concentration condition, and it is a common assumption in similar nonparametric testing problems. When combined with Assumption (5), this assumption imposes mild conditions on the rate at which $\varepsilon_n$ converges to zero (see, for instance, Chernozhukov et al. (2013) for a discussion).

3.4 Motivating the Method of Moments using Re-Centered Moments

In section 3.1, we introduced our estimator as a simple modification applied to a method of moments estimator. Within that section, we claimed the use of the method of moments objective function avoided the non-standard features encountered when using the log-likelihood. To reinforce this, we provide a heuristic illustration of the violation of some of the conditions in section 3.3 when using the (negative) log-likelihood as the criterion function. To simplify the illustration, we temporarily
assume a parametric model with \( f_0(v) = f(v, \theta_0) \) for some \( \theta_0 \in \Theta \) and we assume the existence of a sufficient number of moments in order to invoke the relevant CLT and LLN.

To begin, let \( L_n(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \log(g(b_i, \theta, p)) \) denote the negative log-likelihood function. we can perform and re-arrange the analogous Taylor expansion in equation (10) in section 3.2 to get

\[
\hat{\theta}_{ml} = \left( \arg \min_{\theta \in \hat{\Theta}} \left\| \sqrt{n}(\theta - \theta_0) - H_n^{-1} S_n \right\|_{H_n}^2 \right) + r_n
\]

where

\[
S_n = \sqrt{n} \frac{\partial L_n(\theta_0)}{\partial \theta}, \quad H_n = \frac{\partial^2 L_n(\theta_0)}{\partial \theta \partial \theta'}
\]

and \( r_n \) denotes the difference between the minimizer of the quadratic approximation and the minimizer of \( L_n(\theta) \). From the context of the assumptions in the previous section, there are three issues with expression (14) which all originate from the parameter-dependent support problem.

The most immediate issue which prohibits us from using the log-likelihood in the preceding analysis is the non-zero expected value of the score function. In contrast to the standard likelihood problems, in section 2.2 we show the expected value of the score function conditional on \( p \) auction participants is given by the known function \( \mu(\theta, p) \) which is non-zero in the auction model we consider. Had we minimized the negative of the log-likelihood to obtain the restricted estimator, Assumption 2 requires the mean-zero asymptotic normality of \( H_n^{-1} S_n \) where

\[
S_n = \sqrt{n} \frac{\partial L_n(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log(g(b_i, \theta_0, p))}{\partial \theta}.
\]

By adding and subtracting \( \mu(\theta_0, p) \) to the previous expression we get

\[
S_n = \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{\partial \log(g(b_i, \theta_0, p))}{\partial \theta} - \mu(\theta_0, p) \right) \right) + \sqrt{n} \mu(\theta_0, p)
\]

where the first term in the above expression is a \( \sqrt{n} \) times a sum of mean-zero random variables. While we can use the CLT to approximate the first term in the expression above by a mean-zero normal random variable, the second term is a sequence of non-stochastic terms diverging to infinity. As a consequence, Assumption 2 cannot hold as \( S_n \) is not stochastically bounded if we use the negative log-likelihood as a criterion function for obtaining the restricted estimator. This issue with the log-likelihood is the primary consideration which motivates the use of the moment-based estimator with re-centered moments.
The last two issues involve potential violations of Assumptions 1 and 6 from section 3.3. In standard log-likelihood problems we would show the positive definiteness of the matrix $H_n$ by using the LLN to argue $H_n$ is close to its expected value, say $H$. If we can freely interchange the order of integration and differentiation, $H$ is equal to the information matrix and assuming positive-definiteness of the information matrix is a commonly invoked identification condition. Due to the parameter-dependent support, this argument is no longer valid as we cannot freely interchange the order of integration and differentiation so information equality fails to hold. As a consequence, it is no longer clear that positive-definiteness of $J_n$ in Assumption 6 is a mild assumption. Moreover, if $H_n$ is not positive-definite it may be difficult or impossible to provide conditions under which $r_n$ is asymptotically small (see Assumption 1) as the second-order Taylor expansion then corresponds to an indefinite quadratic form which may have no global extrema on $\hat{\Theta}_R$.

4 Low-Level Conditions for a Simple Auction Model

In this section, we introduce low-level sufficient conditions for the conditions presented in section 3.3. We focus on assumptions for the nonparametric environment. In section 4.1 we discuss the sequence of sieve spaces we use for estimation, and in section 4.2 we state the low-level conditions.

4.1 The Sieve Space

Let $\Theta \subseteq \mathbb{R}^{K_n}$. For nonparametric models, we take $\theta_0 \in \mathbb{R}^{K_n}$ such that $f_0(v) \approx f(v, \theta_0)$, where our low-level conditions then impose restrictions on how slowly/quickly $K_n$ is allowed to grow with $n$.\footnote{We have suppressed the dependence of $\theta_0$ on $n$ for notational convenience.}

For an arbitrary element $\theta \in \Theta$, the associated density is given by

$$f(v, \theta) = (P(v)' \theta)^2 \phi(v)$$

where $\phi(v)$ is a density function supported on $[\underline{v}, \bar{v}]$ which is specified by the researcher and $P(v)$ is a $K_n$-dimensional vector of $\phi(v)$-orthonormalized basis functions. For example, $\phi(v)$ could be the uniform distribution on $[\underline{v}, \bar{v}]$ or a truncated parametric distributions (such as the log-normal). When $\phi(v)$ corresponds to a (truncated) parameterized distribution with parameter $\beta$, we denote this by $\phi(v, \beta)$. The form of $f(v, \theta)$ is common for sieve density estimators as $f(v, \theta)$ is always non-negative. Additionally, the use of $\phi(v)$—orthonormal basis functions $P(v)$ implies $f(v, \theta)$ integrates to one if and only if $\|\theta\| = 1$.\footnote{We have suppressed the dependence of $\theta_0$ on $n$ for notational convenience.}
4.2 Low-Level Sufficient Conditions

In this section, we provide the low-level sufficient conditions for Assumptions (1) - (4) appearing in section 3.3. For the remaining conditions, Assumptions (5) and (7) depend on the particular choice of test-statistic, and Assumption (6) corresponds to the nonparametric analogue of the identification condition for parametric method of moments estimators. We assume the data is generated according to the simple auction model specified in section 2.1 with the true density of valuations given by \( f_0(v) \). Let \( \Theta \) be the sieve-space described in section 4.1. We now get the following result

**Theorem 2.** Let \( \zeta_n \equiv \sup_{v \in [v, \bar{v}]} \| P(v) \| \) and \( \mathcal{P} \) be the class of distributions satisfying the following assumptions. There exists a constant \( 0 < C < \infty \) not depending on \( n \) such that

A.1) For all \( \theta \in \Theta \) and \( v \in [v, \bar{v}] \)

\[
\frac{1}{C} < f_0(v), f(v, \theta), \phi(v) < C,
\]

A.2) there exists \( \theta_0 \in \Theta \) such that for some \( \gamma > 0 \)

\[
\sup_{v \in [v, \bar{v}]} |f(v, \theta_0) - f_0(v)| \leq CK_n^{-\gamma} \quad \text{and} \quad b(\theta_0, p) \geq \bar{b}_0(p)
\]

where \( \bar{b}_0(p) = \bar{v} - \int_{\bar{v}}^{v} F_0(t)p^{-1}dt \) denotes the largest bid rationalized by \( f_0(v) \) and

A.3) the following rate conditions hold

\[
\frac{K_n^2 \zeta_n^4}{n} = o_p(\varepsilon_n), \quad \frac{K_n^4 \zeta_n^2}{n} = o_p(\varepsilon_n), \quad nK_n^{-2\gamma} \zeta_n^2 = o_p(\varepsilon_n).
\]

Assumptions (A.1) - (A.3) imply Assumptions (1) - (4) in section 3.3.

**Remark 1:** The proof of the theorem is included in the online supplemental appendix. Assumption (A.1) bounds the true density away from zero and away from infinity and is a common assumption used in the literature on auctions (see for instance assumption A2 in GPV (2000) and assumption 1.e in Marmer and Shneyerov (2012)). Assumption (A.2) specifies the rate at which the bias of the sieve converges to zero where \( \gamma \) is controlled by the number of continuous derivatives of \( f_0(v) \). Furthermore, this assumption states that the approximating sequence of \( \theta_0 \) approaches \( f_0(v) \) from “inside” the restricted set. Lastly, (A.3) is a set of abstract rate conditions. The first two rates limit the growth of the sieve and ensure the approximation errors from the estimated
covariance and the approximate normality shrink sufficiently fast. The last rate condition specifies a lower bound on the growth-rate of the sieve and is used to ensure the bias of the sieve is shrinking sufficiently fast.

**Remark 2:** If $\zeta_n \leq CK_n$, as is the case when $P(v)$ is taken to be the polynomials, splines, or any bounded set of functions, our rate conditions simplify to require $K_n^6$ and $nK_n^{-2(\gamma - 1)}$ both converge to zero. While the first condition is identical to the analogous rate appearing in Newey (1997), the anologue of the second condition found in Newey (1997) is $nK_n^{-2\gamma}$ converges to zero so our second rate is slightly stronger. However, as illustrated in GPV (2000) to achieve the same optimal rate of convergence for the simple-auction model as in standard density estimation with $R$ continuous derivatives, one must assume the distribution of valuations is $R+1$ times continuously differentiable. Our rate conditions exhibit a similar feature as $\gamma$ is given by the number of (bounded) continuous derivatives of $f_0(v)$. As a result, we must assume an additional order of continuous differentiability to get the same rate conditions as in Newey (1997) for nonparametric regression.

5 Monte Carlo Simulations

In this section, we present the results of a Monte Carlo simulation to demonstrate the finite sample properties of our estimator. We apply our estimator to data generated from the following model. We draw independent valuations $\{v_i\}_{i=1}^n$ from a a truncated gamma-distribution with parameters $(5, 1)$ with truncation occurring at known constants $[2, 10]$. Figure 2 displays the shape of the distribution $f_0(v)$. We then draw valuations for $L = \{100, 200, 300, 400\}$ auctions each containing $p = 5$ auction participants resulting in the total sample sizes $n = \{500, 1,000, 1,500, 2,000\}$. Letting $F_0(v)$ denote the cumulative distribution function of $f_0(v)$, we then generate bids from the valuations $\{v_i\}_{i=1}^n$ according to

$$b_i = v_i - \frac{\int_{v_i}^{v_i} F_0(t)^{p-1} dt}{F_0(v_i)^{p-1}}.$$ 

As practitioners, we only observe the sample $\{(b_i, p_i)\}_{i=1}^L$, and our interest lies in estimating and forming confidence bands for three functionals of interest: 1) the density of valuations at the true median $f(x_0, \theta)$, 2) the expected revenue of the auction $\psi_{rev}(\theta)$ and 3) the optimal reserve price $r_{opt}(\theta)$. Importantly, we do not assume the data was generated from the class of truncated gamma distributions, and we estimate the model nonparametrically.

For these simulations, we use the sieve-class described in section 4.1 with $\phi(v)$ as the log-normal density with parameters $\beta = (\beta_\mu, \beta_\sigma)$. We then obtain $P(v)$ by applying the Gram–Schmidt
orthonormalization process to the set of cubic splines with $K_n$-partitions (i.e. $K_n + 1$ knots and $K_n + 2$ overall parameters). As a practical note, we obtain the parameters $(\beta_\mu, \beta_\sigma)$ and the knot-locations for the splines by the following procedure. First, given the sample $\{b_{il}\}_{i=1}^{p} \{p_l\}_{l=1}^{L}$, we obtain the (parametric) maximum likelihood estimate for $\beta = (\mu, \sigma)$, say $\hat{\beta} = (\hat{\mu}, \hat{\sigma})$. Then, for $j \in \{0, ..., K_n\}$, we obtain the $j$th knot point as $\Phi^{-1}\left(\frac{j}{K_n}, \hat{\beta}\right)$, where $\Phi^{-1}(\cdot, \beta)$ denotes the inverse cdf of $\phi(\cdot, \beta)$. We then obtain the maximum likelihood estimator $\hat{\theta}_{\text{ml}}$ using the point $\theta = [1, 0, ..., 0]'$ as a starting point for optimizer which corresponds to the maximum-likelihood estimator in the parametric submodel $\phi(v, \beta)$. After obtaining the maximum likelihood estimator, denoted $\hat{\theta}_{\text{ml}}$, we use this as a starting-point to find the method of moments estimator, denoted $\hat{\theta}_{\text{mm}}$, and then use the $\hat{\theta}_{\text{mm}}$ to form our proposed estimator $\hat{\theta}_n$ according to equation (7). When obtaining $\hat{\theta}_{\text{ml}}$ and $\hat{\theta}_{\text{mm}}$, we enforce the constraint $b(\theta, p) \geq \max\{b_{il}\}$ in the optimization.6

Table (1) reports the root-mean-squared error (RMSE) of the three indicated functionals using three different plug-in estimates: 1) our estimator $\hat{\theta}_n$, 2) the method of moments estimator $\hat{\theta}_{\text{mm}}$ and 3) the maximum log-likelihood estimator $\hat{\theta}_{\text{ml}}$. Due to space considerations, we only report results for $K_n = 3$ knots and the results for the other values of $K_n$ are qualitatively similarly. In all cases, the bias of the estimates are quite small so the RMSE is driven primarily by the standard deviation of the estimators. A striking feature of this table is that the mean-squared-error of the three plug-in estimators are all of similar magnitudes. As $\hat{\theta}_{\text{mm}}$ and $\hat{\theta}_{\text{ml}}$ are restricted estimators and our estimator is not restricted to lie in $\hat{\theta}_n$, we may expect that the RMSE of our estimator is higher than the MM and ML estimators as the latter are estimates formed using a smaller parameter space. Despite this fact, the relative variance of our proposed estimator is comparable to the variances of the ML and MM estimators. In fact, our estimator has at most 5% higher variance than either the ML or MM estimators for the three functionals across all specifications. Consequently, this suggests the variance cost of using the normally distributed estimator $\hat{\theta}_n$ as opposed to the non-normal MM and ML estimators is small.

Table (2) reports the empirical coverage probabilities of a standard two-sided $t$-test of the indicated functionals using three different plug-in estimates: 1) our estimator $\hat{\theta}_n$, 2) the method of moments estimator $\hat{\theta}_{\text{mm}}$ and 3) the maximum log-likelihood estimator $\hat{\theta}_{\text{ml}}$. Due to space considerations, we only report results for $K_n = 3$ knots and the results for the other values of $K_n$ are qualitatively similarly. In all cases, the bias of the estimates are quite small so the RMSE is driven primarily by the standard deviation of the estimators. A striking feature of this table is that the mean-squared-error of the three plug-in estimators are all of similar magnitudes. As $\hat{\theta}_{\text{mm}}$ and $\hat{\theta}_{\text{ml}}$ are restricted estimators and our estimator is not restricted to lie in $\hat{\theta}_n$, we may expect that the RMSE of our estimator is higher than the MM and ML estimators as the latter are estimates formed using a smaller parameter space. Despite this fact, the relative variance of our proposed estimator is comparable to the variances of the ML and MM estimators. In fact, our estimator has at most 5% higher variance than either the ML or MM estimators for the three functionals across all specifications. Consequently, this suggests the variance cost of using the normally distributed estimator $\hat{\theta}_n$ as opposed to the non-normal MM and ML estimators is small.

5To denote the set of splines on $[v, \tau]$ with $K_n$ partitions let the knots be given by $v = t_0 < t_1 < ... < t_{K_n-1} < t_{K_n} = \tau$. Then, the set of degree $m$ splines is given by the set

$\{1, x, ..., x^m, \max\{x - t_1, 0\}^m, ..., \max\{x - t_{K_n-1}, 0\}^m\}$

6Strictly speaking, this is not necessary. However, the optimization routine we use (fmincon) tends to work better if we prohibit the optimizer from searching over values of $\theta$ for which the criterion function is not defined.
cated functional at the 5% level. Our proposed confidence bands perform well and have empirical coverage close to the nominal size. The low coverage for $K_n = 1$ with larger sample sizes is due to the bias in the sieve-approximation. This bias disappears when more series terms are added as coverage returns to 95%. Additionally, the high coverage for $K_n = 4$ for small sample sizes indicates the importance of using a slowly growing sieve in controlling the size of the test.

Table (3) reports the performance of our plug-in proposed estimator as measured in mean-squared-error as it compares to alternatives in the literature on first-price auctions. The two alternatives we consider are the estimator of GPV (2000) and the estimator proposed in Marmer and Shneyerov (2012). In comparison to these alternatives, our estimator appears to perform well with mean-squared errors significantly smaller than those found using the alternative estimators over the range of $K_n$ values we used.\footnote{All results in Table (3) which report the results of the alternative estimators uses our implementation of their estimators. To mitigate the chance of substantial coding errors, we used our implementation of their estimators to replicate the simulation studies appearing in their papers. Using this method, we were able to quantitatively replicate the results in the papers Guerre et al. (2000), Marmer and Shneyerov (2012) and Ma et al. (2018). The results and codes used in these replications are available upon request.}

Table (4) reports the empirical coverage results of our estimator as compared to the available alternatives for the indicated functionals where we have used the bandwidth choices suggested in those papers. For the density of valuations, while all three methods apply and appear to control size, the confidence sets using our method are substantially more narrow than the two alternatives. In particular, for sample sizes large enough to obtain proper size control, their proposed confidence bands are on average over four times larger than our confidence bands. For the optimal reserve price, only our paper and Ma et al. (2018) provide an asymptotically valid testing procedure. To produce their confidence sets, we invert the test of the optimal reserve price described in the appendix of their paper. Our test appears to have better size control for smaller sample sizes while producing confidence sets which are, on average, half the size of their confidence sets. Moreover, in contrast to our procedure, the confidence sets for $r_{opt}$ in Marmer and Shneyerov (2012) are not guaranteed to be an interval as they can be arbitrary subsets of the support of valuations.\footnote{In fact, over all of our Monte Carlo simulations, none of the confidence sets were intervals using this procedure.} Lastly, our method is the only available approach in the literature which can produce confidence sets for the expected revenue using the optimal reserve price.
6 Application to US Timber Auctions

As an application of our method, we focus on timber auctions conducted by the United States Forest Service. The US Forest Service administers approximately 193 million acres of federally-owned land and is responsible for maintaining the majority of federal timber lands within the United States. To maintain the health of the forests they administer and provide the sawmills within the United States with a sufficient amount of timber, the USFS frequently sells the right to harvest timber on tracts of land using the first-price auction format. Within these auctions, sawmills and logging companies submit bids in order to purchase the right to harvest the timber on a tract of land over a specified length of time. Before announcing an auction, the Forest Service conducts a “cruise” of a tract of land to estimate the quantity and quality of timber on the tract of land. For each species found in substantial quantities on the land, the cruise report contains information on the quantity of timber (usually measured in thousand-board-feet\(^9\)) and an appraisal on the expected market value of the timber based on the anticipated quality of the wood.\(^{10}\) Additionally, the cruise also contains an estimate of the cost of harvesting and manufacturing the timber into a final product.\(^{11}\)

The primary participants in the auctions are operators of sawmills and logging companies. Both types of bidders are specialized in the type of tracts/timber with which they work. For example, logging companies may specialize in clear-cutting or thinning operations and may also specialize according to the type of logging/harvesting which can be performed on a particular tract of land.\(^{12}\) Further, sawmills are highly specialized not just in the particular species of timber they use but they may also specialize in particular cuts of wood and production of finished products. Lastly, sawmills differ substantially in their production costs. Specialized equipment and trade secrets enable sawmills to extract different amounts of usable wood product from a fixed amount of timber, which is referred to as the “overrun” rate of the mill. According to Baldwin et al. (1997),

---

\(^9\) A board-foot is the quantity of wood contained in a one-foot square piece of wood which is one-inch thick.

\(^{10}\) Numerous factors contribute to the quality of the wood, such as the height/thickness of the tree, the number of knots and other defects in the wood, etc. Typically, the wood is assigned a grade based on the quality and use of the lumber it can produce. For example, construction-grade lumber can be considerably more valuable than logs which are deemed utility/pulp grade.

\(^{11}\) The data also contains estimates of the cost of temporary road construction/maintenance which the USFS expects the winning firm to incur to extract the timber from the tract. To avoid complicated policies regarding the (partial) reimbursement of construction costs, we focus on auctions which do not require road construction or maintenance.

\(^{12}\) For example, on relatively flat tracts the most efficient technique may be using ground-based extraction equipment/techniques whereas these may be unavailable/impractical on tracts of land with steep slopes or in narrow valleys where cable-yarding/skyline equipment/techniques are more appropriate.
the overrun of a sawmill is “a closely guarded secret,” and constitutes a substantial source of private information for bidders.

We focus on sealed-bid and scaled-sale auctions from 1981-1993.\textsuperscript{13} For the auctions considered, agents observe the cruise report and submit a sealed-bid for each species appearing in the cruise report. Using the quantity estimates in the cruise report, the Forest Service then uses the species-specific bid rates submitted by the bidders to compute the total value of the bid. The contract is then awarded to the agent that submitted the largest bid in the auction.\textsuperscript{14} As the winning agency removes timber from the forest, they pay the Forest Service their bid rate on the actual amount of each species they harvest from the land. As bidders in these auctions face less risk due to the uncertainty over the volume of timber on a tract in comparison with lump-sum auctions, participants in these auctions are less likely to conduct their own cruise of the land. Beacuase of the high degree of specialization of bidders in conjunction with the use of scaled-sale format, the assumption of independent private valuations is plausible for the auctions under consideration. Several other papers in the literature (see Paarsch (1997), Haile (2001), Haile and Tamer (2003), Haile et al. (2006) among others) make similar assumptions.

An implicit assumption we make when taking the auction model to the data is that the sample consists of repeated, independent auctions of a homogeneous item which does not vary from auction to auction. For timber auctions, there is considerable heterogeneity in the characteristics of the tract of land as well as the timber being auctioned. Specifically, tracts of land may differ substantially in the acreage, density of the timber, distance from local mills, terrain of the tract, composition of the species of trees, etc. For what follows, we assume that the only observables which directly affect an agent’s bid are the logging/manufacturing cost estimates and the final market-value appraisal of the

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{13} We focus on auctions with contract dates after May 4th, 1981 due to a policy change in the auctions starting at that date. Prior to this date, winning agencies were able to re-sell their rights to a third party. As documented in Haile (2001), this introduces the potential for a common-value component of the auction in which bidders also speculate about the re-sale value of the contract.
\item \textsuperscript{14} For the analysis here, we implicitly assume agents do not engage in bid-skewing. Bid-skewing refers to the practice of altering the per-species bid rate while keeping the overall bid unchanged in order to take advantage of ex post realizations of quantities. For example, if a firm believes the forest service has underestimated/overestimated timber of a particular species, they may increase/decrease their bid for this species (keeping their overall bid quantity unchanged) to still win the bid but pay a smaller amount of fees for harvesting the timber. See Athey and Levin (2001) for more details.
\end{itemize}
\end{footnotesize}
timber provided by the Forest Service.\textsuperscript{15} To address this heterogeneity, we adjust bids in our sample to reflect the amount an agent is willing to pay to extract and manufacture a fixed unit of median-quality timber.\textsuperscript{16} Specifically, we adjust a bid by taking the total bid submitted by an agent plus the total estimated logging and manufacturing costs divided by the number of quality-standardized board-feet of lumber on the tract. We construct the measure of quality-standardized board-feet for an auction by adjusting the reported amount of timber for each species to the amount of median-value timber needed to keep the total market value of the two quantities identical. As an example, suppose the cruise report contains 100,000 board-feet of low-quality Oak with an appraised market value of $0.3/bf. If the median price of all observations among all species is $0.4/bf, we say the number of quality-standardized board-feet on the tract is

\[
100,000 \, \text{BF}_{\text{Oak}} \times \frac{0.3}{\text{BF}_{\text{med. tree}}} = 75,000 \, \text{BF}_{\text{med. tree}}
\]

After this adjustment, a bid within our data set reflects the total anticipated cost a bidder expects to pay to produce a single board-foot of median-quality timber.

Thus far, we have assumed agents know the number of participants within the auction when submitting their bid. As the auctions we consider here are sealed-bid auctions in which agents mail their bid to a sale officer, it is unlikely that agents know the exact number of participants who are also submitting bids. A more palatable assumption, therefore, is to assume that the number of bidding participants is random and that an agent knows the distribution the number of participants.\textsuperscript{17} Let \( \pi_p \) an agent’s belief that there will be \( p \) participants (including the agent) who

\textsuperscript{15}In combination, these two pieces of information jointly specify the total estimated cost for removing the timber as well as the expected market value of the timber once harvested. We assume all other observable characteristics only influence bidding behavior through these two channels.

\textsuperscript{16}An alternative approach to control this heterogeneity would be to condition on auction-specific covariates. In this approach, one assumes valuations are additively separable into an auction-specific and individual-specific component. For example, such a specification could be

\[
v_{it} = X_{it}\beta + \eta_i
\]

where \( X_{it} \) reflects the auction-specific covariates and \( \eta_i \) reflects the idiosyncratic, private, independent valuation for the object. It can be shown that this additive structure (when combined with risk-neutrality) one can first regress observed bids on the auction-specific covariates and use the orthogonal residuals as the bids resulting from the private independent valuations. A challenge in using this approach is correctly accounting for the error in estimating \( \beta \) in the construction of confidence sets for functionals of the distribution of \( \eta_i \). This is an avenue for future research.

\textsuperscript{17}This auction environment has been considered previously by Harstad et al. (1997). Song (2006) provides a result on the identification and consistent estimation of this auction format.
submit bids in the auction and assume the number of potential participants is bounded. It can be shown that the equilibrium bidding function is given by

\[ b(v, \theta, \pi) = \sum_p w_p(v, \theta, \pi) b(v, \theta, p) \]

where \( w_p(v, \theta, \pi) = \pi_p F(v, \theta)^{\theta^{-1}} / \sum \pi_p F(v, \theta)^{\theta^{-1}} \) and \( b(v, \theta, p) \) is the so-called contingent bidding function given in equation (1). For what follows we assume a known value of \( \pi \).

We can use arguments in equation (2) to derive the bid-density as

\[ g(b, \theta, \pi) = f(\eta(b, \theta, \pi), \theta) \eta'(b, \theta, \pi), \]

where \( \eta'(b, \theta, \pi) \) is the derivative of the derivative of the inverse bid function with respect to the bids. Furthermore, as \( w_p(\theta, \theta, \pi) = \pi_p \), the support of bids depends on the true parameters as

\[ \bar{b}(\theta, \pi) = \sum_p \pi_p \bar{b}(\theta, p). \]

Similar to the arguments in section 2.2 we can apply Leibniz integral rule to get an identical expression as equation (6) using \( \bar{b}(\theta, \pi) \) in place of \( \bar{b}(\theta, p) \). Similarly, we can get the new value of \( \mu(\theta, \pi) \) in the first equality in equation (6). With the value \( \mu(\theta, \pi) \), all arguments of section 3 and the general theory in section 3.3 remain valid. In the empirical analysis that follows, we use the empirical distribution of \( p \) in the auction data to construct a pre-estimate of \( \pi \). We take the value of \( \pi \) as fixed, so our confidence bands only incorporate uncertainty in the estimation of \( \theta \).

Our final sample of bids consists of 4,458 bids from 1,129 auctions. Figure (3) contains the empirical density of the submitted bids and Table (5) contains distribution of auctions across the number of potential players. To arrive at this final data set, we only kept data which passed several cleaning checks. As a preliminary step, we only kept data from auctions using the sealed-bid format for which at least two bids were received and all volumes of timber were measured in thousands of board-feet with no sale restrictions. As an additional level of cleaning, we only kept data where the number of acres of the tract was non-zero, all bids were greater than 1/20 times but less than

\[ \text{All of the arguments which follow still hold under an unknown } \pi. \text{ The only change is that in deriving the function } \mu(\theta, \pi) \text{ we must differentiate with respect to both parameters } \theta \text{ and } \pi, \text{ which makes the notation cumbersome. Our current estimates use a pre-estimated plug-in estimate for } \pi, \text{ but we are actively working on relaxing this to jointly estimate } (\theta, \pi). \]

\[ \text{This is given as the reciprocal of the derivative of equation (16) with respect to } v. \]

\[ \text{We are currently working on extending the code to jointly estimate } (\pi, \theta). \]

\[ \text{Occasionally, the USFS will hold auctions in which companies with more than 500 employees are prohibited from bidding. These are referred to as SBA-set asides, and are commonly excluded in empirical studies of this industry.} \]
20 times the total appraised value of the tract, at least half of the volume of timber had a positive appraised value, no temporary road construction/maintenance was required, the appraised value per board-foot of all species was within eight standard deviations of the average appraised value for that species, and all species on the tract were species for which we had at least one-hundred recorded valuations in the data.\textsuperscript{22}

To estimate the density of valuations, we used quadratic splines with five knot points. The lower-bound of valuations, \( v \), was set as the smallest observed bid while the upper-bound of the valuations was chosen using a nested-maximum likelihood procedure.\textsuperscript{23} Lastly, as there is a potential for measurement error in the submitted bids, we follow the method of Aarts et al. (2007) to estimate the right endpoint of the support of the distribution of bids as the \( 1 - \beta_n \) quantile of the empirical distribution of bids for a vanish sequence \( \beta_n \).\textsuperscript{24,25} Figure (5) displays the estimated density of valuations along with the upper and lower bands of the (point-wise) 90\% confidence band. Figure 5 displays the implied distribution of bids along with the histogram of observed bids. Our estimate for the density of valuations implies a plug-in estimate for the optimal reserve price of $0.253 per standardized board foot with a resulting confidence band of [$0.2387, $0.2673] using the two-sided t-statistic the 5\% level (all prices measured in 2012 dollars).

As the stated objective of the US Forest Service involves maintaining the health and productivity of the forest under its care while also providing a sufficient supply of timber to the nation’s sawmills for a fair market price, it is unlikely that the revenue generated from the auctions is sole basis for setting the reserve price.\textsuperscript{26} Instead, we assume the Forest Service will wish to set a reserve price not only on the basis of the revenue it may generate but also taking into account the probability in which no bidder is able to submit a bid above the reserve price and the timber rights go unsold.

\textsuperscript{22}The appendix contains more detailed information on the data cleaning.

\textsuperscript{23}For a value of the upper bound, \( \bar{v} \), we can compute the maximum likelihood estimator \( \hat{\theta}_{\text{mle}}(\bar{v}) \). We then take the value of \( \bar{v} \) for which \( L_n(\hat{\theta}_{\text{mle}}(\bar{v})) \) was the largest. When forming confidence sets, we treat \( \bar{v} \) as a known parameter.

\textsuperscript{24}As an example of the potential for measurement error, before cleaning the data, 49 bids per standardized-board feet exceeded $10.00, which is approximately 60 times the median bid.

\textsuperscript{25}The results in Aarts et al. (2007) deal with estimation of the support of the distribution \( X^* \) when one only observes contaminated data \( X = X^* + u \). They demonstrate maximum order statistic of \( X \) is generally not consistent for the upper bound of the support of \( X^* \) if \( u \) is a non-degenerate random variable. Furthermore, they show that if the rate of convergence of the estimate for the distribution of \( X \) is \( \alpha_n \), then the \( (1 - \beta_n) \) quantile of the observed data is a consistent estimate of the support of \( X^* \) when \( \beta_n \to 0 \) and \( \frac{\alpha_n}{\beta_n} \to 0 \). We use this estimator with \( \beta_n = \frac{1}{n^{1/3}} \).

\textsuperscript{26}This is evidenced by fact that the Forest Service has reportedly accepted bids for timber-removal contracts which do not even generate enough revenue to cover the department’s administration fees for monitoring/conducting the auction.
The latter consideration may be important for the Forest Service when selling timber rights for tracts of land which are overgrown or contain a large number of dead trees and may represent a substantial fire hazard for the surrounding forest. Without knowing the welfare function of the USFS, however, it is impossible to provide an estimate of the welfare maximizing reserve price. As a compromise, we provide point-estimates of both the expected revenue of the auction and the probability the tract will be sold as a function of the reserve price, \( r \). This allows us to illustrate the tradeoffs policy makers face when setting the reserve price as it impacts the expected revenue and the probability of no-sale. Figure 6 contains both of these estimated functions as well as (joint) uniform 95% confidence bands obtained using the two-sided t-statistics. This figure shows that in order to increase the expected revenue of the auction slightly, the USFS would have to face a substantial reduction in the sale probability. Consequently, if non-revenue considerations enter the welfare function, it is possible the US Forest Service could optimally set reserve prices far lower than the revenue-optimizing reserve price.

Using our estimator and proposed inference procedure, operators of a firm or policymakers in a government agency can easily estimate and form confidence sets for the reserve price which best achieves the objectives of the organization. To illustrate this, let \( W(r, \theta) \) denote the utility of setting a reserve price \( r \) when the density of valuations is given by \( f(v, \theta) \). We could obtain a plug-in estimate for the welfare optimizing reserve price, say \( r(\hat{\theta}_n) \), by maximizing \( W(r, \hat{\theta}_n) \). Assuming the welfare function is continuously differentiable and the welfare-maximizing reserve price is in the interior of the valuation space, we could use our method to obtain the asymptotic normality of \( \sqrt{n}(r(\hat{\theta}_n) - r(\theta_0)) \). Using our result, a policymaker can obtain a simple 95% confidence set as

\[
\left[ r(\hat{\theta}_n) - 1.96 \frac{\hat{\sigma}_r}{\sqrt{n}}, r(\hat{\theta}_n) + 1.96 \frac{\hat{\sigma}_r}{\sqrt{n}} \right]
\]

where \( \hat{\sigma}_r^2 = \frac{\partial r(\hat{\theta}_n)}{\partial \theta'} \hat{\Sigma}_n \frac{\partial r(\hat{\theta}_n)}{\partial \theta} \).

No other method in the literature on first-price auctions can produce an analogous confidence band.

7 Conclusion

In this paper, we proposed a new estimator for the distribution of valuations in first-price auctions with independent and private valuations. Our estimator is constructed as a modified method of moments estimator which avoids the issues one encounters when trying to analyze the properties of the maximum likelihood estimator. As our estimator avoids the non-standard features of the auctions model, our estimator has a normal limiting distribution which allows simple construction of confidence sets for a wide array of features of interest.
Although we focus on a simple first-price auction model, our method applies more generally to construct asymptotically normally distributed estimators from restricted estimators when the criterion function may only be defined on a strict subset of the parameter space. As such, our method may be useful for other, more elaborate first-price auction models such as independent private valuation models with binding reserve prices, risk adverse bidders, auction-specific covariates, etc. or to models which potentially affiliated or common values. Within each of these models, the support of the bid distribution depends non-trivially upon the parameters of the valuation distribution so the log-likelihood is only defined for a subset of the parameters. As result, our method may be useful for constructing asymptotically normally distributed estimators in these models. The only requirement for constructing the estimator is that the density and support of the bid distribution are closed-form functions of the parameters of the valuation distribution, as is often the case for these models. If the high-level conditions proposed in this paper apply, the resulting estimator will be asymptotically normally distributed. It is an open question whether one can find low-level sufficient conditions for these assumptions for particular variants of the auction model.
References


## 8 Appendix

### 8.1 Appendix : Tables

<table>
<thead>
<tr>
<th>Functional</th>
<th>n</th>
<th>Root Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Our Estimator</td>
</tr>
<tr>
<td>Density at a Point</td>
<td>500</td>
<td>0.0159</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>0.0093</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>0.0079</td>
</tr>
<tr>
<td>Optimal Reserve Price</td>
<td>500</td>
<td>0.1158</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>0.0661</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>0.0574</td>
</tr>
<tr>
<td>Expected Revenue</td>
<td>500</td>
<td>0.0669</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>0.0472</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>0.0360</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>0.0328</td>
</tr>
</tbody>
</table>

Table 1: This table presents the square-root of the mean squared error (rmse) for the indicated functionals for the three indicated estimators. As the bias of the three estimators is small, the rmse is driven primarily by the standard deviation of the estimators. Therefore, this suggests that our estimator has only a slightly larger variance than the method of moments or maximum likelihood estimators. All results are for $K_n = 3$ partitions, but the qualitative results are present for the other values of $K_n$ used in the simulation.
Empirical Coverage : Our Proposed Estimator

<table>
<thead>
<tr>
<th>$K_n$</th>
<th>$n$</th>
<th>$f_0(x_0)$</th>
<th>$r_{opt}$</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500</td>
<td>94.7%</td>
<td>93.9%</td>
<td>92.3%</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>92.4%</td>
<td>93.6%</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>90.6%</td>
<td>94.4%</td>
<td>93.6%</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>87.0%</td>
<td>91.2%</td>
<td>93.1%</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>94.4%</td>
<td>92.5%</td>
<td>93.8%</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>93.8%</td>
<td>94.9%</td>
<td>94.0%</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>93.8%</td>
<td>94.9%</td>
<td>94.1%</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>95.2%</td>
<td>95.3%</td>
<td>95.4%</td>
</tr>
<tr>
<td>3</td>
<td>500</td>
<td>95.3%</td>
<td>94.2%</td>
<td>94.8%</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>95.2%</td>
<td>94.1%</td>
<td>93.8%</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>94.6%</td>
<td>94.1%</td>
<td>96.2%</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>94.6%</td>
<td>93.2%</td>
<td>94.9%</td>
</tr>
<tr>
<td>4</td>
<td>500</td>
<td>97.2%</td>
<td>96.6%</td>
<td>97.6%</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>93.8%</td>
<td>94.2%</td>
<td>95.5%</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>94.7%</td>
<td>95.8%</td>
<td>95.0%</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>93.2%</td>
<td>95.6%</td>
<td>97.0%</td>
</tr>
</tbody>
</table>

Table 2: This table records the empirical coverage rates for the indicated functionals. All coverage results are for the two-sided $t$-test conducted at the 5% level.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Density at a Point</td>
<td>500</td>
<td>0.0084</td>
<td>0.0117</td>
<td>0.0159</td>
<td>0.0364</td>
<td>0.0364</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>0.0065</td>
<td>0.0066</td>
<td>0.0111</td>
<td>0.0297</td>
<td>0.0311</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>0.0057</td>
<td>0.0066</td>
<td>0.0093</td>
<td>0.0250</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>0.0053</td>
<td>0.0055</td>
<td>0.0079</td>
<td>0.0230</td>
<td>0.0251</td>
</tr>
<tr>
<td>Optimal Reserve Price</td>
<td>500</td>
<td>0.0942</td>
<td>0.1166</td>
<td>0.1158</td>
<td>0.2178</td>
<td>0.2761</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>0.0674</td>
<td>0.0638</td>
<td>0.0820</td>
<td>0.1681</td>
<td>0.2282</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>0.0549</td>
<td>0.0638</td>
<td>0.0661</td>
<td>0.1499</td>
<td>0.1980</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>0.0510</td>
<td>0.0549</td>
<td>0.0574</td>
<td>0.1341</td>
<td>0.1752</td>
</tr>
<tr>
<td>Expected Revenue</td>
<td>500</td>
<td>0.0641</td>
<td>0.0668</td>
<td>0.0669</td>
<td>0.1623</td>
<td>0.3101</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>0.0439</td>
<td>0.0384</td>
<td>0.0472</td>
<td>0.1317</td>
<td>0.1615</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>0.0355</td>
<td>0.0386</td>
<td>0.0360</td>
<td>0.1205</td>
<td>0.1049</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>0.0321</td>
<td>0.0322</td>
<td>0.0328</td>
<td>0.1102</td>
<td>0.0721</td>
</tr>
</tbody>
</table>

Table 3: This table compares the square root of the mean-squared error of our estimator to the two alternatives proposed in Marmer and Shneyerov (2012) and Guerre et al. (2000).
Table 4: This table compares our estimator to the results of the inference procedures of Marmer and Shneyerov (2012) and Ma et al. (2018) where the latter are applicable using the tuning parameters suggested in those papers. To avoid unfair comparisons (i.e. those with our method using a small number of series terms to get artificially small confidence sets), our results are reported cubic splines with $K_n = 4$ partitions. Moreover, we augment the results with a sample-size of $n = 4,000$, which is (approximately) the sample size used in the MC simulations of the two competing papers. While all approaches appear to control size well for large sample sizes, our approach is more broadly applicable, delivers smaller confidence sets when multiple approaches apply and appears to deliver better finite-sample coverage for smaller sample sizes.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Coverage</td>
<td>Avg CS Size</td>
<td>Coverage</td>
</tr>
<tr>
<td>Density at a Point</td>
<td>500</td>
<td>97.2%</td>
<td>0.054</td>
<td>90.2%</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>93.8%</td>
<td>0.038</td>
<td>93.5%</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>94.7%</td>
<td>0.031</td>
<td>93.4%</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>93.2%</td>
<td>0.027</td>
<td>94.7%</td>
</tr>
<tr>
<td></td>
<td>4,000</td>
<td>94.9%</td>
<td>0.019</td>
<td>97.6%</td>
</tr>
<tr>
<td>Optimal Reserve</td>
<td>500</td>
<td>96.6%</td>
<td>0.430</td>
<td>83.3%</td>
</tr>
<tr>
<td>Price</td>
<td>1,000</td>
<td>94.2%</td>
<td>0.301</td>
<td>86.9%</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>95.8%</td>
<td>0.247</td>
<td>87.6%</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>95.6%</td>
<td>0.215</td>
<td>88.9%</td>
</tr>
<tr>
<td></td>
<td>4,000</td>
<td>94.3%</td>
<td>0.152</td>
<td>91.4%</td>
</tr>
<tr>
<td>Expected Revenue</td>
<td>500</td>
<td>97.6%</td>
<td>0.258</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>1,000</td>
<td>95.5%</td>
<td>0.180</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>1,500</td>
<td>95.0%</td>
<td>0.147</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>2,000</td>
<td>97.0%</td>
<td>0.127</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>4,000</td>
<td>95.4%</td>
<td>0.089</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 5: This table contains the breakdown of the number of auctions by the number of participants $p$. We use these numbers to compute the plug-in estimate for $\pi$ described in section 6.

<table>
<thead>
<tr>
<th>$p$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>324</td>
<td>246</td>
<td>189</td>
<td>127</td>
<td>105</td>
<td>75</td>
<td>41</td>
<td>22</td>
<td>1,129</td>
</tr>
</tbody>
</table>
8.2 Appendix : Figures

Figure 1: This figure plots the parametric distribution of valuations discussed in section 2.2.

Figure 2: This figure contains the true density of the Monte Carlo simulation presented in section 3.
Figure 3: This figure illustrates a kernel-smoothed estimate of the density of bids for the application to USFS timber auctions.

Figure 4: This figure illustrates a kernel-smoothed estimate of the density of all appraised prices for all species of timber.
Figure 5: This figure contains our point-estimate for the density of valuations using quadratic splines with four knots. The dashed lines correspond to 95% uniform confidence bands for the true density of valuations.

Figure 6: This figure illustrates the trade-off between higher revenues and a lower probability of selling timber as the reserve price is increased from no reserve price to the revenue-maximizing reserve price. The decreasing function (and the associated dashed lines) correspond to the point estimate and uniform confidence band for the probability of conducting a sale as a function of the reserve price. The increasing function (and the associated dashed lines) correspond to the estimate and confidence bands for the expected revenue of the auction as a function of the sale price. The dashed lines correspond to (joint) uniform 95% confidence bands. This figure shows that, while increasing the reserve price increases the expected revenue, the increase in revenue is small (only 4% increase over the entire domain) whereas there is a precipitous drop in the sale probability as the reserve price is increased.
9 Technical Appendix

9.1 Proof of theorem

Let $Z_n$ denote the normal random variable described in assumption $[2]$. In the first part of the proof of the theorem, we establish that $\|\sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n\| = o_p(\varepsilon_n)$. Our proof is based on a slight extension of a similar proof appearing in Ketz (2018). We start by re-arranging an expression for our proposed estimator

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\hat{\theta}_{mm} - \theta_0) - \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \sqrt{n} \frac{\partial Q_n(\hat{\theta}_{mm})}{\partial \theta} \]

\[
= -\left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \sqrt{n} \left( \frac{\partial Q_n(\hat{\theta}_0)}{\partial \theta} \right) + \left( I_{d_{\theta}} - \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \frac{\partial^2 Q_n(\theta^+)}{\partial \theta^2} \right) \sqrt{n}(\hat{\theta}_{mm} - \theta_0)
\]

\[
= \sqrt{n}(\theta_n - \theta_0) + \left( \left( \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right)^{-1} - \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \right) \sqrt{n} \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} \right)
\]

\[
+ \left( I_{d_{\theta}} - \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \frac{\partial^2 Q_n(\theta^+)}{\partial \theta^2} \right) \sqrt{n}(\hat{\theta}_{mm} - \theta_0)
\]

where the second line follows from a mean-value expansion of $\frac{\partial Q_n(\theta)}{\partial \theta}$ around $\theta_0$ and $\theta^+$ is a point with $\|\theta^+ - \theta_0\| \leq \|\hat{\theta}_{mm} - \theta_0\|$. The third line uses the definition $\sqrt{n}(\theta_n - \theta_0) = \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} \right)^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}$. Therefore, by the triangle inequality and the properties of the spectral norm, we can re-arrange the above equation to get

\[
\left\| \sqrt{n}(\hat{\theta}_n - \theta_0) - \sqrt{n}(\theta_n - \theta_0) \right\| \leq \left\| \left( \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right)^{-1} - \left( \frac{\partial^2 Q_n(\hat{\theta}_0)}{\partial \theta^2} \right)^{-1} \right\| \left\| \sqrt{n} \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} \right) \right\|
\]

\[
+ \left\| I_{d_{\theta}} - \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \frac{\partial^2 Q_n(\theta^+)}{\partial \theta^2} \right\| \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) \right\|
\]

\[
\leq \lambda_{max} \left( \left( \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right)^{-1} \right) \lambda_{max} \left( \left( \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right)^{-1} \right) \left\| \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} - \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right\| \left\| \sqrt{n} \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} \right) \right\|
\]

\[
+ \lambda_{max} \left( \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \right) \left\| \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} - \frac{\partial^2 Q_n(\theta^+)}{\partial \theta^2} \right\| \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) \right\|
\]

\[
\leq \lambda_{max} \left( \left( \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right)^{-1} \right) L_n \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) \right\| \left\| \sqrt{n} \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} \right) \right\|
\]

\[
+ 2\lambda_{max} \left( \left( \frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta^2} \right)^{-1} \right) L_n \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) \right\|^2
\]

(17)\[
= O_p \left( \frac{L_n \sqrt{K_n}}{\sqrt{n}} \right) \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) \right\| + O_p \left( \frac{L_n}{\sqrt{n}} \right) \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) \right\|^2
\]


\[
\left\| \sqrt{n} \left( \frac{\partial Q_n(\theta_0)}{\partial \theta} \right) \right\| \leq \lambda_{min} \left( \left( \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right)^{-1} \right) \left\| \frac{\partial^2 Q_n(\theta_0)}{\partial \theta^2} \right\| \left\| \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \right\| \leq O_p(1) \lambda_{min} \left( \Sigma_n \right)^{-\frac{1}{2}} \left\| \Sigma_n^{-\frac{1}{2}} Z_n \right\| + o_p(1) = O_p(\sqrt{K_n})
\]

where the second inequality uses assumption $[2]$ and $[6]$. Furthermore, by assumptions $[1]$ and $[6]$, we have

\[
\left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) - \sqrt{n}(\theta_n - \theta_0) \right\| \leq \lambda_{min}(J_n)^{-\frac{1}{2}} \left\| \sqrt{n}(\hat{\theta}_{mm} - \theta_0) - \sqrt{n}(\theta_n - \theta_0) \right\| J_n
\]

\[
\leq \lambda_{min}(J_n)^{-\frac{1}{2}} \left\| \sqrt{n}(\theta_n - \theta_0) \right\| J_n
\]

\[
\leq \lambda_{min}(J_n)^{-\frac{1}{2}} \left\| \Sigma_n \right\| J_n + \lambda_{min}(J_n)^{-\frac{1}{2}} o_p(\varepsilon_n)
\]

\[
\leq \lambda_{max}(J_n)^{-\frac{1}{2}} \lambda_{min}(J_n)^{-\frac{1}{2}} \lambda_{min}(\Sigma_n)^{-\frac{1}{2}} \left\| \Sigma_n^{-\frac{1}{2}} Z_n \right\| + o_p(\varepsilon_n)
\]

\[
= O_p(\sqrt{K_n})
\]
where the second inequality uses assumption (1) and the last line uses assumption (6), assumption (2) and Markov’s inequality. Using the triangle inequality then gives

\[ \| \sqrt{n}(\hat{\theta}_{\text{mm}} - \theta_0) \| \leq \| \sqrt{n}(\hat{\theta}_{\text{mm}} - \sqrt{n}(\theta_n - \theta_0)) \| + \| \sqrt{n}(\theta_n - \theta_0) \| = O_p(\sqrt{K_n}) \]

where the equality uses Markov’s inequality. Plugging this bound into equation (17) and applying the rate condition on \( L_n \) in assumption (4) gives

\[ \| \sqrt{n}(\hat{\theta}_n - \theta_0) - \sqrt{n}(\theta_n - \theta_0) \| = o_p(\epsilon_n). \]

Therefore, by the triangle inequality, we get

\[ \| \sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n \| \leq \| \sqrt{n}(\hat{\theta}_n - \theta_0) - \sqrt{n}(\theta_n - \theta_0) \| + \| \sqrt{n}(\theta_n - \theta_0) - Z_n \| = o_p(\epsilon_n). \]

In the remainder of the proof, we show that this approximation error combined with the error from estimating the covariance matrix and critical values have a negligible impact on the distribution of the test-statistic. This proof follows a similar but substantially simplified version of a proof appearing in Freyberger and Reeves (2018). By assumptions (2) and (5) and the triangle-inequality we have

\[ \left| T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) - T(Z_n, \Sigma_n) \right| \leq \left| T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) - T(Z_n, \hat{\Sigma}_n) \right| + \left| T(Z_n, \hat{\Sigma}_n) - T(Z_n, \Sigma_n) \right| \]

\[ \leq C\left\| \sqrt{n}(\hat{\theta}_n - \theta_0) - Z_n \right\| + C\left\| \Sigma_n \right\| \left\| \hat{\Sigma}_n - \Sigma_n \right\| \]

\[ = o_p(\epsilon_n) \]

(20)

where the last line follows from Markov’s inequality, assumption (3) and the observation that

\[ \left\| Z_n \right\| = \left\| \Sigma_n^{\frac{1}{2}} \Sigma_n^{-\frac{1}{2}} Z_n \right\| \leq \lambda_{\text{max}}(\Sigma_n)^{\frac{1}{2}} \left\| N(0, I_{K_n}) \right\| = \lambda_{\text{max}}(\Sigma_n)^{\frac{1}{2}} \sqrt{K_n} \]

Letting \( A_n \) denote the event that \( \left| T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) - T(Z_n, \Sigma_n) \right| < \frac{1}{2}\epsilon_n \), the above shows \( P(A_n^c) = o(1) \). By basic probability rules

\[ P\left( T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) \leq c_{1-\alpha,n}(\hat{\Sigma}_n) \right) \leq P\left( T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) \leq c_{1-\alpha,n}(\hat{\Sigma}_n), A_n \right) + P\left( A_n^c \right) \]

\[ \leq P\left( T(Z_n, \Sigma_n) \leq c_{1-\alpha,n}(\Sigma_n) + \frac{1}{2}\epsilon_n \right) + o(1) \]

(21)

Similarly,

\[ P\left( T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) \leq c_{1-\alpha,n}(\hat{\Sigma}_n) \right) \geq P\left( T\left(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n\right) \leq c_{1-\alpha,n}(\hat{\Sigma}_n), A_n \epsilon_n \right) \]

\[ \geq P\left( T(Z_n, \Sigma_n) \leq c_{1-\alpha,n}(\Sigma_n) + \frac{1}{2}\epsilon_n \right) + P\left( A_n \epsilon_n \right) \]

\[ \geq P\left( T(Z_n, \Sigma_n) \leq c_{1-\alpha,n}(\Sigma_n) - \frac{1}{2}\epsilon_n \right) - P\left( A_n^c \right) \]

\[ = P\left( T(Z_n, \Sigma_n) \leq c_{1-\alpha,n}(\Sigma_n) - \frac{1}{2}\epsilon_n \right) - o(1) \]

(22)

where the second inequality follows as the event in the third line implies the event in the second line and the final inequality comes from basic probability identities.\(^{27}\) The previous two displays provide upper and lower bounds on the coverage probability of the proposed test. Specifically, these equations quantify the impact of the normal approximation on the coverage probability of the test. Note, if \( \epsilon_n = 0 \) and \( \hat{\Sigma}_n = \Sigma_n \), the previous displays would provide

\[^{27}\text{Note, the inequality uses the fact that for any two measurable events } A, B, \text{ we have } P(A) = P(A \cap B) + P(A \cap B^c) \text{ which implies } P(A \cap B) = P(A) - P(A \cap B^c) \geq P(A) - P(B^c) \]
the result. In the next part of the proof, we establish that the approximation error in using \( \hat{\Sigma}_n \) is asymptotically small. Specifically, we show

\[
(23) \quad c_{1-a,n}(\hat{\Sigma}_n) \geq c_{1-a-\delta_q,n}(\Sigma_n) - \frac{1}{2} \varepsilon_n
\]

and

\[
(24) \quad c_{1-a,n}(\hat{\Sigma}_n) \leq c_{1-a+\delta_q,n}(\Sigma_n) + \frac{1}{2} \varepsilon_n
\]

To establish the first result, first notice by assumptions (5) and (3) and Markov’s inequality,

\[
\|T(Z_n, \hat{\Sigma}_n) - T(Z_n, \Sigma_n)\| \leq C\|Z_n\||\hat{\Sigma}_n - \Sigma_n\| = o_p(\varepsilon_n).
\]

Therefore, letting \( B_n \) denote the event that \( ||T(Z_n, \hat{\Sigma}_n) - T(Z_n, \Sigma_n)|| < \frac{1}{2} \varepsilon_n \), the previous display shows \( P(B_n^C) = o(1) \). Therefore, by the definition of the critical value and basic probability rules

\[
1 - \alpha = P\left(T(Z_n, \hat{\Sigma}_n) \leq c_{1-a,n}(\Sigma)\right) \\
\leq P\left(T(Z_n, \Sigma_n) \leq c_{1-a,n}(\Sigma) + \frac{1}{2} \varepsilon_n, B_n\right) + P\left(B_n^C\right) \\
\leq P\left(T(Z_n, \Sigma_n) \leq c_{1-a,n}(\Sigma) + \frac{1}{2} \varepsilon_n\right) + o(1)
\]

where the \( o(1) \) term can be made small uniformly over \( P \in \mathcal{P} \). Specifically, the \( o(1) \) term can be made smaller than \( \delta_q \) uniformly over \( P \in \mathcal{P} \) so that the previous display shows

\[
1 - \alpha - \delta_q \leq P\left(T(Z_n, \Sigma_n) \leq c_{1-a,n}(\Sigma) + \frac{1}{2} \varepsilon_n\right)
\]

so that \( c_{1-a-\delta_q,n}(\Sigma) \leq c_{1-a,n}(\Sigma) + \frac{1}{2} \varepsilon_n \), which establishes the first result. To establish equation (24) we follow an analogous argument to get

\[
1 - \alpha = P\left(T(Z_n, \Sigma_n) \leq c_{1-a,n}(\Sigma_n)\right) \\
\geq P\left(T(Z_n, \hat{\Sigma}_n) + \frac{1}{2} \varepsilon_n \leq c_{1-a,n}(\Sigma_n), B_n\right) \\
\geq P\left(T(Z_n, \hat{\Sigma}_n) \leq c_{1-a,n}(\Sigma_n) - \frac{1}{2} \varepsilon_n\right) - P\left(B_n^C\right) \\
= P\left(T(Z_n, \hat{\Sigma}_n) \leq c_{1-a,n}(\Sigma_n) - \frac{1}{2} \varepsilon_n\right) - o(1)
\]

where we can make the \( o(1) \) term uniformly (over \( P \)) smaller than \( \delta_q \). Hence, we have

\[
1 - \alpha + \delta_q \geq P\left(T(Z_n, \hat{\Sigma}_n) \leq c_{1-a,n}(\Sigma_n) - \frac{1}{2} \varepsilon_n\right)
\]

so that \( c_{1-a+\delta_q}(\hat{\Sigma}) \geq c_{1-a,n}(\Sigma_n) - \frac{1}{2} \varepsilon_n \) which establishes equation (24). Combining equations (21) and (23) gives

\[
P\left(T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1-a,n}(\Sigma_n) + \varepsilon_n\right) \geq P\left(T(Z_n, \Sigma_n) \leq c_{1-a}(\Sigma_n) + \frac{1}{2} \varepsilon_n\right) - o(1) \\
\geq P\left(T(Z_n, \Sigma_n) \leq c_{1-a-\delta_q,n}(\Sigma_n)\right) - o(1) \\
= 1 - \alpha - \delta_q - o(1).
\]

As we can make the \( o(1) \) term arbitrarily small uniformly over \( P \) and the choice of \( \delta_q \) is arbitrary, we have

\[
\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1-a,n}(\Sigma_n) + \varepsilon_n\right) \geq 1 - \alpha
\]

39
which is the first conclusion of the theorem. Additionally, for any $\delta_q$ equations (22) and equation (23) imply

$$P(T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1 - \alpha, n}(\hat{\Sigma}_n)) \geq P(T(Z_n, \Sigma_n) \leq c_{1 - \alpha, n}(\Sigma) - \frac{1}{2} \varepsilon_n)$$

$$\geq P(T(Z_n, \Sigma_n) \leq c_{1 - \alpha - \delta_q, n}(\Sigma) - \varepsilon_n).$$

Similarly, by equations (22) and (24) we have

$$P(T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1 - \alpha, n}(\hat{\Sigma}_n)) \leq P(T(Z_n, \Sigma_n) \leq c_{1 - \alpha, n}(\hat{\Sigma}_n) + \frac{1}{2} \varepsilon_n) + o(1)$$

$$\leq P(T(Z_n, \Sigma_n) \leq c_{1 - \alpha + \delta_q, n}(\Sigma_n) + \varepsilon_n) + o(1)$$

Therefore, if assumption (7) holds, then for each $\delta_q$, we have

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}} \left| P(T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1 - \alpha, n}(\hat{\Sigma})) - (1 - \alpha) \right| \to 0,$$

which establishes the theorem.

9.2 Supplementary Lemmas

Lemma 1. Useful Properties of Eigenvalues

Let $A_n$ and $\hat{A}_n$ be symmetric, positive definite matrices. Then

$$|\lambda_{\text{max}}(A_n) - \lambda_{\text{max}}(\hat{A}_n)| \leq \|A_n - \hat{A}_n\|$$

$$|\lambda_{\text{min}}(A_n) - \lambda_{\text{min}}(\hat{A}_n)| \leq \|A_n - \hat{A}_n\|$$

where $\|A\|$ denotes the Frobenius norm of $A$.

Proof. By the definition of the smallest eigenvalues,

$$\lambda_{\text{min}}(A_n) = \inf_{\|x\|=1} x' A_n x \leq \sup_{\|x\|=1} |x'(A_n - \hat{A}_n)x| \leq \|A_n - \hat{A}_n\| \leq \|A_n - \hat{A}_n\|.$$

Similarly, by the definition of the largest eigenvalues,

$$\lambda_{\text{max}}(A_n) = \sup_{\|x\|=1} x' A_n x \leq \sup_{\|x\|=1} |x'(A_n - \hat{A}_n)x| \leq \|A_n - \hat{A}_n\| \leq \|A_n - \hat{A}_n\|.$$

Lemma 2. Plug-in Second Derivative is Positive Definite

Under assumptions (1), (2), (4), and (6), the eigenvalues of $\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial \theta \partial \theta'}$ are bounded and bounded away from zero.

Proof. To begin, notice

$$\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} = \left( \frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} \right).$$

From assumptions (4) and (6) along with properties of norms we have

$$\|\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}}) - \partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'}\| \leq \frac{L_n}{\sqrt{n}} \|\sqrt{n}(\hat{\theta}_{\text{mm}} - \theta_0)\| \leq \frac{L_n}{\sqrt{n}} C \|\sqrt{n}(\hat{\theta}_{\text{mm}} - \theta_0)\| J_n \leq \frac{L_n}{\sqrt{n}} C \|J_n^{-1} \sqrt{n} D_n\| J_n + \frac{L_n}{\sqrt{n}} o_p(\varepsilon_n)$$

where the last inequality follows from assumption (1). Using properties of norms and assumption (2) we have

$$\|J_n^{-1} \sqrt{n} D_n\| J_n \leq C \|J_n^{-1} \sqrt{n} D_n\| \leq C \|Z_n\| + o_p(\varepsilon_n) \leq C^2 \|\Sigma_{n}^{\frac{1}{2}} Z_n\| + o_p(\varepsilon_n) = O_p(\sqrt{k_n})$$

40
where the last line uses Markov’s inequality and the fact that $\Sigma_n^{1/2}Z_n \sim N(0, I_{K_n})$. Using the rate conditions on $L_n$, we can combine the last several statements to get $\|\frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'}\| = o_p(1)$. As the Frobenius norm is the square-root of the sum of square eigenvalues, the previous display shows the absolute value of the eigenvalues of $\frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'}$ are in any arbitrarily small neighborhood of zero with probability approaching one. With this fact, an application of Weyl’s inequality—which uses the assumed symmetry of the matrices—to expression (25) shows the smallest eigenvalue of $\frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta \partial \theta'}$ is within an arbitrarily small neighborhood of the smallest eigenvalue of $\frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'}$ with probability approaching one. As the latter is strictly larger than zero, the $\frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial \theta \partial \theta'}$ is also strictly larger than zero with probability approaching one.

Lemma 3. Plug-in Estimate of $\Sigma_n$ is Positive Definite
Under assumptions (3) and (6), $\hat{\Sigma}_n$ is symmetric and positive definite with probability approaching one.

Proof. First, notice $\hat{\Sigma}_n = \Sigma_n + (\hat{\Sigma}_n - \Sigma_n)$ where $\|\hat{\Sigma}_n - \Sigma_n\| = o_p(1)$. An application of Weyl’s inequality—which uses the assumed symmetry of $\hat{\Sigma}_n$—implies the smallest/largest eigenvalues of $\hat{\Sigma}_n$ are within a fixed but arbitrarily small neighborhood of the smallest/largest eigenvalues of $\Sigma_n$ with probability approaching one. By choosing a sufficiently small neighborhood, this implies the smallest/largest eigenvalues of $\hat{\Sigma}_n$ are bounded away from zero and bounded away from infinity with probability approaching one, which is the statement of the lemma.