Shape-Restricted Problems in Econometrics

By

Brandon Reeves

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(ECONOMICS)

at the

UNIVERSITY OF WISCONSIN – MADISON

2019

Date of final oral examination: February 11, 2019

The dissertation is approved by the following members of the Final Oral Committee:

Professor J. Freyberger, Assistant Professor, Economics

Professor B. Hansen, Professor, Economics

Professor H. Kang, Assistant Professor, Statistics

Professor J. Porter, Professor, Economics

Professor X. Shi, Associate Professor, Economics

Abstract

Many econometric models restrict the set of parameters consistent with the underlying model theory or observable data. These restrictions may come from *a priori* beliefs about the model–such as monotonicity of demand curves–or from statistical or economic theory–such as non-crossing quantile functions or first-price auction models which restrict the set of observable bids. When these restrictions are binding or close to binding, standard asymptotic theory provides poor approximations to the finite-sample behavior of the restricted estimator. This dissertation concerns the construction of estimators and inference procedures for shape-restricted models.

The first chapter proposes a uniformly valid inference method for a parameter vector satisfying certain shape-restrictions. The method applies generally to a range of finite dimensional and nonparametric problems, such as regressions or instrumental variable estimation, to both kernel or series estimators, and to many shape restrictions. The bands are asymptotically equivalent to standard, unrestricted confidence bands if the true parameter strictly satisfies all shape restrictions, but they can be much smaller if some of the shape restrictions are binding or close to binding. We illustrate these sizable width gains in Monte Carlo simulations and in an empirical application.

The second chapter proposes a general method for constructing asymptotically normally distributed estimators from shape-restricted estimators which applies in both parametric and nonparametric settings. Due to the asymptotically normality, our estimator avoids the non-standard distribution of shape-restricted estimators. Consequently, our resulting confidence sets are easy to obtain and simple to report. As our main application of interest, we provide low-level assumptions under which our method applies to the estimation of first-price auctions with independent, private valuations. In this context, our method provides the first inference result in the literature which allows for the construction of confidence sets for a general class of functions. Simulations suggest our estimator and inference procedure perform well, as our confidence sets have empirical coverage near nominal levels and the mean squared error of our estimator compares favorably against alternative estimators in the literature. We demonstrate the empirical usefulness of our approach in an application to timber auctions conducted by the US Forest Service.

Acknowledgements

This dissertation would not have been possible without the advice and support of numerous faculty members at the University of Wisconsin-Madison. I am indebted to Joachim Freyberger, Bruce Hansen, Jack Porter and Xiaoxia Shi. Through your courses and the discussions we had, you shaped the direction of my research and solidified my passion for econometrics. I was beyond fortunate to have Joachim Freyberger as my advisor for this dissertation. Through generously offering his time and expertise, Joachim was instrumental in the completion of this thesis, and I could not have asked for a better advisor. In addition, I would like to thank Hyunseung Kang for graciously volunteering his time to serve on my dissertation committee.

Throughout my years in graduate school, I would not have been able to keep my sanity if it were not for my friends. Whether it was discussing problem sets or just hanging out at the terrace, I am grateful to have had the support of great friends like Adam, Daniel, Emilio, Gabriel and Moheeb. I was also fortunate to have friends like Joey, Riley, Sam and Shazaer, who offered much-needed distractions from my research when I would encounter roadblocks.

Lastly, I would to thank my family—Ryan, Jenny, Kris and Jay—for their constant support, not just while in graduate school, but throughout all of my life. In particular, I would like to thank my parents for their years of unending encouragement. I would not be where I am today without everything you all have given me.

List of Figures

List of Tables

Contents

Chapter 1

Inference under Shape Restrictions[1]

1.1 Introduction

Researchers can often use either parametric or nonparametric methods to estimate the parameters of a model. Parametric estimators have favorable properties, such as good finite sample precision and fast rates of convergence, and it is usually straightforward to use them for inference. However, parametric models are often misspecified. Specifically, economic theory rarely implies a particular functional form, such as a linear or quadratic demand function, and conclusions drawn from an incorrect parametric model can be misleading. Nonparametric methods, on the other hand, do not impose strong functional form assumptions, but as a consequence, confidence intervals obtained from them are often much wider

In this paper we explore shape restrictions to restrict the class of functions but without imposing arbitrary parametric assumptions. Shape restrictions are often reasonable assumptions, such as assuming that the return to eduction is positive, and they can be implied by economic theory. For example, demand functions are generally monotonically decreasing in

[1]**This chapter is joint work with Joachim Freyberger**

prices, cost functions are monotonically increasing, homogeneous of degree 1, and concave in input prices, Engel curves of normal goods are monotonically increasing, economies of scale yield subadditive average cost functions, and utility functions of risk averse agents are concave. Additionally, statistical theory can imply shape restrictions, such as noncrossing conditional quantile curves. There is a long history of estimation under shape restrictions in econometrics and statistics and obtaining shape restricted estimators is simple in many settings. Moreover, shape restricted estimators can have much better finite sample properties, such as lower mean squared errors, compared to unrestricted estimators.

Using shape restrictions for inference is much more complicated than simply obtaining a restricted estimator. The main reason is that the distribution of the restricted estimator depends on where the shape restrictions bind, which is unknown a priori. In this paper we propose a uniformly valid inference method for an unknown function or parameter vector satisfying certain shape restrictions, which can be used to test hypotheses and to obtain confidence sets. The method applies very generally, namely to a wide range of finite dimensional and nonparametric problems, such as regressions or instrumental variable estimation, to both kernel or series estimators, and to many different shape restrictions. Our confidence sets are well suited to be reported along with shape restricted estimates, because they are build around restricted estimators and eliminate regions of the parameter space that are inconsistent with the shape restrictions.

One major application of our inference method is to construct uniform confidence bands for a function. Such a band consists of a lower bound function and an upper bound function such that the true function is between them with at least a pre-specified probability. These bands are useful to summarize statistical uncertainty and they allow the reader to easily assess statistical accuracy and perform various hypothesis tests about the function without access to the data. Our confidence bands have desirable properties. In particular, they always include the shape restricted estimator of the function and are therefore never empty. Moreover, they are asymptotically equivalent to standard unrestricted confidence bands if the true function strictly satisfies all shape restrictions (e.g. if the true function is strictly increasing but the shape restriction is that it is weakly increasing). However, if for the true function some of the shape restrictions are binding or close to binding, our confidence bands are generally much smaller. The decrease in the width reflects the increased precision of the constrained estimator. Finally, the proposed method provides uniformly valid inference over a large class of distributions, which in particular implies that the confidence bands do not suffer from undercoverage if some of the shape restrictions are close to binding. These cases are empirically relevant. For example, demand functions are likely to be strictly decreasing, but nonparametric estimates are often not monotone, suggesting that the demand function is close to constant for some prices.[2] Our method applies very generally. For example, our paper is the first to provide such inference results for the nonparametric instrumental variables (NPIV) model under general shape constraints. The method can also be used to obtain confidence intervals for functionals, such as average derivatives, with similar properties.

Similar to many other nonstandard inference problems, instead of trying to obtain confidence sets directly from the asymptotic distribution of the estimator, our inference procedure is based on test inversion.[3] This means that we start by testing the null hypothesis that the true parameter vector θ_0 is equal to some fixed value $\bar{\theta}$. In series estimation θ_0 represents the coefficients in the series approximation of a function and θ_0 can therefore grow in dimension as the sample size increases. The major advantage of the test inversion approach is that under the null hypothesis we know exactly which of the shape restrictions are binding or close to binding. Therefore, under the null hypothesis, we can approximate the distribution of the estimator in

[2] Analogously to many other papers, closeness to the boundary is relative to the sample size.

[3] Other nonstandard inference settings include autoregressive models (e.g. Mikusheva (2007)), weak identification (e.g. Andrews and Cheng (2012)), and partial identification (e.g. Andrews and Soares (2010)).

large samples and we can decide whether or not we reject the null hypothesis. The confidence set for θ_0 consists of all values for which the null hypothesis is not rejected.

To obtain uniform confidence bands or confidence sets for other functions of θ_0, such as average derivatives, we project onto the confidence set for θ_0 (see Section 1.2 for a simple illustration). We choose the test statistic in a way that our confidence sets are asymptotically equivalent to standard unrestricted confidence sets if θ_0 is sufficiently in the interior of the parameter space. Thus, in this case, the confidence sets have the right coverage asymptotically. If some of the shape restrictions are binding or close to binding, our inference procedure will generally be conservative due to the projection. However, in these cases we also obtain very sizable width gains compared to a standard unrestricted confidence set. Furthermore, due to test inversion and projections, our inference method can be computationally demanding. We provide details on the computational costs and compare them with alternative approaches in Section 1.6.1. We also briefly describe a method recently suggested by Kaido et al. (2016) in a computationally similar problem in the moment inequality literature, which also applies to our framework, and can reduce these costs considerably.

In Monte Carlo simulations we construct uniform confidence bands in a series regression framework and in the NPIV model under a monotonicity constraint. In the NPIV model the gains of using shape restrictions are generally much higher. For example, we show that with a fourth order polynomial approximation of the true function, the average width gains can be up to 73%, depending on the slope of the true function. We also obtain large widths gains for confidence intervals for the average derivative of the function. Finally, in an empirical application, we estimate demand functions for gasoline, subject to the functions being weakly decreasing, and we provide uniform confidence bands build around restricted estimates with monotone upper and lower bound functions. In this setting, the width gains from using these shape restrictions are between 25% and 45%.

We now explain how our paper fits into the related literature. There is a vast literature on estimation under shape restrictions going back to Hildreth (1954) and Brunk (1955) who suggest estimators under concavity and monotonicity restrictions, respectively. Other related work includes, among many others, Mukerjee (1988), Dierckx (1980), Ramsay (1988), Mammen (1991a), Mammen (1991b), Mammen and Thomas-Agnan (1999), Hall and Huang (2001), Haag et al. (2009), Du et al. (2013), and Wang and Shen (2013). See also Delecroix and Thomas-Agnan (2000) and Henderson and Parmeter (2009) for additional references. Many of the early papers focus on implementation issues and subsequent papers discuss rates of convergence of shape restricted estimators. Many inference results, such as those by Mammen (1991b), Groeneboom et al. (2001), Dette et al. (2006), Birke and Dette (2007), and Pal and Woodroofe (2007) are for points of the function where the shape restrictions do not bind. It is also well known that a shape restricted estimator has a nonstandard distribution if the shape restrictions bind; see for example Wright (1981) and Geyer (1994). Freyberger and Horowitz (2015) provide inference methods in a partially identified NPIV model under shape restrictions with discrete regressors and instruments. Empirical applications include Matzkin (1994), Lewbel (1995), Ait-Sahalia and Duarte (2003), Beresteanu (2005), and Blundell, Horowitz, and Parey (2012, 2017). There is also an interesting literature on risk bounds (e.g. Zhang (2002), Chatterjee et al. (2015), Chetverikov and Wilhelm (2017)) showing, among others, that a restricted estimator can have a faster rate of convergence when the true function is close to the boundary. In addition, there is a large, less related literature on testing shape restrictions. See also Chetverikov et al. (2018) for a recent review article.

There are several existing methods which can be used to obtain uniform confidence bands under shape restrictions. First, there are a variety of existing bands, some of which are tailored to specific shape restrictions, such as the ones in Dümbgen (1998, 2003), which have the feature that they can be empty with positive probability. As a very simple example, one could intersect a standard unrestricted band with all functions satisfying the shape restrictions. While one could interpret an empty band as evidence against the shape restrictions, these bands can also be arbitrarily small, and the width might thus not adequately reflect the finite sample uncertainty. More formally, these bands do not satisfy the "reasonableness" property of Müller and Norets (2016).

The method of Dümbgen (2003) only applies to a regression model with fixed regressors and normally distributed errors, but he shows that his bands adapt to the smoothness of the unknown function.[4] focus on confidence intervals for the function evaluated at a point in the normal regression model and they show that the intervals adapt to each individual function under monotonicity and convexity constraints. Bellec (2016) constructs polyhedron type confidence regions for the entire conditional mean vector in the normal regression model with shape restrictions, but without using any smoothness assumptions, and he shows that they adapt the dimension of the smallest face of the polyhedron. One could use these results to construct uniform confidence bands or confidence intervals for functionals by projecting onto these confidence sets, but the resulting bands or intervals would be very conservative. Our method applies much more generally and covers, among other, the NPIV model under general shape constraints, but investigating analogous adaptivity results is out of scope of the current paper. However, we briefly compare the two approaches in simulations. We then also show that our bands are on average much narrower than simple monotonized bands. The second possibility is to use the rearrangement idea of Chernozhukov et al. (2009), which works with monotonicity restrictions and has recently been extended to some other shape restrictions by Chen et al. (2018). While this method is very easy to implement, rearranging a band to monotonize it does not change its average width. Finally, in a kernel regression framework with very general constraints, one could use a two step procedure by Horowitz and Lee (2017).

[4]Cai et al. (2013)

In the first step, they estimate the points where the shape restrictions bind. In the second step, they estimate the function under equality constraints and hence, they obtain an asymptotically normally distributed estimator, which they can use to obtain uniform confidence bands. While their approach is computationally much simpler than ours, their main result leads to bands which can suffer from under-coverage if some of the shape restrictions are close to binding. They also suggest using a bias correction term to improve the finite sample coverage probability, but they do not provide any theoretical results for this method. To the best of our knowledge, our method is the first that yields uniform confidence bands, which are uniformly valid, yield width reductions when the shape restrictions are binding or close to binding, and are never empty.

An additional closely related paper, which does not consider uniform confidence bands and therefore does not fit into any of the categories above, is Chernozhukov et al. (2015). They develop a general testing procedure in a conditional moments setting, which can be used to test shape restrictions and to obtain confidence regions for functionals under shape restrictions. They allow for partial identification, while we assume point identification, but we study a general setup, which includes for example maximum likelihood estimation and conditional moments models. Even though there is some overlap in the settings where both methods apply, their approach is conceptually very different to ours. Similar to many other papers in the partial identification literature, to obtain confidence regions for functionals, they use a test inversion procedure, which jointly tests certain features of the model. In particular, they invert a joint test of the null hypothesis that the shape restrictions hold and that a functional takes a particular value. Consequently, the resulting confidence regions represent both uncertainty about the value of the functional and the shape restrictions, these sets can be empty (which could be interpreted as evidence against the shape restrictions), and they can be arbitrarily small. Contrarily, we impose the shape restrictions and test a null hypothesis about the parameter vector only. Thus, we treat these restrictions and other assumptions of the model, such as moment conditions, symmetrically. Our resulting confidence sets therefore represent uncertainty about the parameter vector only. We illustrate these conceptual differences as well as the computational costs in Section 1.6, where we consider confidence intervals for an average derivative.

Finally, our paper builds on previous work on inference in nonstandard problems, most importantly the papers of Andrews (1999, 2001) on estimation and testing when a parameter is on the boundary of the parameter space. The main difference of our paper to Andrews' work is that we allow testing for a growing parameter vector while Andrews considers a vector of a fixed dimension. Moreover, we show that our inference method is uniformly valid when the parameters can be either at the boundary, close to the boundary, or away from the boundary. We also use different test statistics because we invert them to obtain confidence bands. Thus, while the general approach is similar, the details of the arguments are very different. Ketz (2018) has a similar setup as Andrews but allows for certain parameter sequences that are close to the boundary under non-negativity constraints.

Outline: We start by illustrating the most important features of our approach in a very simple example. Section 1.3 discusses a general setting, including high level assumptions for uniformly valid inference. Sections 1.4 and 1.5 provide low level conditions in a regression framework (for both series and kernel estimation) and the NPIV model, respectively. The remaining sections contain Monte Carlo simulations, the empirical application, and a conclusion. Proofs of the results from Sections 1.4 and 1.5, computational details, and additional simulation results are in a supplementary appendix with section numbers S.1, S.2, etc..

Notation: For any matrix A, $\|A\|$ denotes the Frobenius norm. For any square matrix A, $\|A\|_S = \sup_{\|x\|=1} \|Ax\|$ denotes the spectral norm. For a positive semi-definite matrix Ω and a vector a let $\|a\|_\Omega = \sqrt{a'\Omega a}$. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and the largest eigenvalue of a symmetric square matrix A. For a sequence of random variables X_n and a class of distributions $\mathcal{P}$ we say $X_n = o_p(\varepsilon_n)$ uniformly over $P \in \mathcal{P}$ if $\sup_{P\in\mathcal{P}} P\left(|X_n| \geq \delta\varepsilon_n\right) \to 0$ for any $\delta > 0$. We say $X_n = O_p(\varepsilon_n)$ uniformly over $P \in \mathcal{P}$ if for any $\delta > 0$ there are M_δ and N_δ such that $\sup_{P\in\mathcal{P}} P\left(|X_n| \geq M_\delta\varepsilon_n\right) \leq \delta$ for all $n \geq N_\delta$.

1.2 Illustrative example

We now illustrate the main features of our method in a very simple example. We then explain how these ideas can easily be generalized before introducing the general setup with formal assumptions in Section 1.3. Suppose that $X \sim N(\theta_0, I_{2\times 2})$ and that we observe a random sample $\{X_i\}_{i=1}^n$ of X. Denote the sample average by $\bar{X}$. We are interested in estimating θ_0 under the assumption that $\theta_{0,1} \leq \theta_{0,2}$. An unrestricted estimator of θ_0, denoted by $\hat{\theta}_{ur}$, is

$$\hat{\theta}_{ur} = \underset{\theta\in\mathbb{R}^2}{\arg\min}\ (\theta_1 - \bar{X}_1)^2 + (\theta_2 - \bar{X}_2)^2.$$

Hence $\hat{\theta}_{ur} = \bar{X}$. Analogously, a restricted estimator is

$$\begin{aligned} \hat{\theta}_r &= \underset{\theta\in\mathbb{R}^2:\ \theta_1\leq\theta_2}{\arg\min}\ (\theta_1 - \bar{X}_1)^2 + (\theta_2 - \bar{X}_2)^2 \\ &= \underset{\theta\in\mathbb{R}^2:\ \theta_1\leq\theta_2}{\arg\min}\ \|\theta - \hat{\theta}_{ur}\|^2, \end{aligned}$$

which implies that $\hat{\theta}_r$ is simply the projecting of $\hat{\theta}_{ur}$ onto $\{\theta \in \mathbb{R}^2 :\ \theta_1 \leq \theta_2\}$. Adding and subtracting θ_0 and multiplying by $\sqrt{n}$ then yields

$$\hat{\theta}_r = \underset{\theta\in\mathbb{R}^2:\ \theta_1-\theta_2\leq 0}{\arg\min}\ \|\sqrt{n}(\theta - \theta_0) - \sqrt{n}(\hat{\theta}_{ur} - \theta_0)\|^2.$$

Let $\lambda = \sqrt{n}(\theta - \theta_0)$. From a change of variables it then follows that

$$\sqrt{n}(\hat{\theta}_r - \theta_0) = \underset{\lambda\in\mathbb{R}^2:\ \lambda_1-\lambda_2\leq\sqrt{n}(\theta_{0,2}-\theta_{0,1})}{\arg\min}\ \|\lambda - \sqrt{n}(\hat{\theta}_{ur} - \theta_0)\|^2.$$

Let $Z \sim N(0, I_{2\times 2})$. Since $\sqrt{n}(\hat{\theta}_{ur} - \theta_0) \sim N(0, I_{2\times 2})$ we get

$$\sqrt{n}(\hat{\theta}_r - \theta_0) \overset{d}{=} \underset{\lambda\in\mathbb{R}^2:\ \lambda_1-\lambda_2\leq\sqrt{n}(\theta_{0,2}-\theta_{0,1})}{\arg\min}\ \|\lambda - Z\|^2,$$

where $\overset{d}{=}$ means that the random variables on the left and right side have the same distribution. Notice that while the distribution of $\sqrt{n}(\hat{\theta}_{ur}-\theta_0)$ does not depend on θ_0 and n, the distribution of $\sqrt{n}(\hat{\theta}_r-\theta_0)$ depends on $\sqrt{n}(\theta_{0,2}-\theta_{0,1})$, which measures how close θ_0 is to the boundary of the parameter space relative to n. We denote a random variable which has the same distribution as $\sqrt{n}(\hat{\theta}_r-\theta_0)$ by $Z_n(\theta_0)$. As an example, suppose that $\theta_{0,1}=\theta_{0,2}$. Then $Z_n(\theta_0)$ is the projection of Z onto the set $\{z \in \mathbb{R}^2 : z_1 \leq z_2\}$.

A 95% confidence region for θ_0 using the unrestricted estimator can be constructed by finding the constant c_{ur} such that

$$P(\max\{|Z_1|,|Z_2|\} \leq c_{ur}) = 0.95.$$

It then follows immediately that

$$P\left(\hat{\theta}_{ur,1}-\frac{c_{ur}}{\sqrt{n}} \leq \theta_{0,1} \leq \hat{\theta}_{ur,1}+\frac{c_{ur}}{\sqrt{n}} \text{ and } \hat{\theta}_{ur,2}-\frac{c_{ur}}{\sqrt{n}} \leq \theta_{0,2} \leq \hat{\theta}_{ur,2}+\frac{c_{ur}}{\sqrt{n}}\right) = 0.95.$$

Thus

$$CI_{ur} = \left\{\theta \in \mathbb{R}^2 : \hat{\theta}_{ur,1}-\frac{c_{ur}}{\sqrt{n}} \leq \theta_1 \leq \hat{\theta}_{ur,1}+\frac{c_{ur}}{\sqrt{n}} \text{ and } \hat{\theta}_{ur,2}-\frac{c_{ur}}{\sqrt{n}} \leq \theta_2 \leq \hat{\theta}_{ur,2}+\frac{c_{ur}}{\sqrt{n}}\right\}$$

is a 95% confidence set for θ_0. While there are many different 95% confidence regions for θ_0, rectangular regions are particularly easy to report (especially in larger dimensions), because one only has to report the extreme points of each coordinate.

Similarly, now looking at the restricted estimator, for each $\theta \in \mathbb{R}^2$ let $c_{r,n}(\theta)$ be such that

$$P(\max\{|Z_{n,1}(\theta)|,|Z_{n,2}(\theta)|\} \leq c_{r,n}(\theta)) = 0.95$$

and define CI_r as

$$\left\{\theta \in \mathbb{R}^2 : \theta_1 \leq \theta_2,\, \hat{\theta}_{r,j}-\frac{c_{r,n}(\theta)}{\sqrt{n}} \leq \theta_j \leq \hat{\theta}_{r,j}+\frac{c_{r,n}(\theta)}{\sqrt{n}},\ j \in \{1,2\}\right\}.$$

Again, by construction $P(\theta_0 \in CI_r) = 0.95$.

Figure 1 in Appendix (A.1) illustrates the relation between c_{ur} and $c_{r,n}(\theta)$. The first panel shows a random sample of Z. The dashed square contains all $z \in \mathbb{R}^2$ such that $\max\{|z_1|, |z_2|\} \leq c_{ur}$. The second panel displays the corresponding random sample of $Z_n(\theta_0)$ when $\sqrt{n}(\theta_{0,2} - \theta_{0,1}) = 0$, which is simply the projection of Z onto the set $\{z \in \mathbb{R}^2 : z_1 \leq z_2\}$. In particular, for each realization z we have $z_n(\theta_0) = z$ if $z_1 \leq z_2$ and $z_n(\theta_0) = 0.5(z_1 + z_2, z_1 + z_2)'$ if $z_1 > z_2$. Therefore, if $\max\{|z_1|, |z_2|\} \leq c_{ur}$, then also $\max\{|z_{n,1}(\theta_0)|, |z_{n,2}(\theta_0)|\} \leq c_{ur}$, which immediately implies that $c_{r,n}(\theta_0) \leq c_{ur}$. The solid square contains all $z \in \mathbb{R}^2$ such that $\max\{|z_1|, |z_2|\} \leq c_{r,n}(\theta_0)$, which is strictly inside the dashed square. The third and fourth panel show a similar situations with $\sqrt{n}(\theta_{0,2} - \theta_{0,1}) = 1$ and $\sqrt{n}(\theta_{0,2} - \theta_{0,1}) = 5$, respectively. As $\sqrt{n}(\theta_{0,2} - \theta_{0,1})$ increases, the percentage projected onto the solid line decreases and thus $c_{r,n}(\theta_0)$ gets closer to c_{ur}. Moreover, once $\sqrt{n}(\theta_{0,2} - \theta_{0,1})$ is large enough, $c_{r,n}(\theta_0) = c_{ur}$.

Figure 2 in Appendix (A.1) shows the resulting confidence regions for θ_0 when $n = 100$ for specific realizations of $\hat{\theta}_{ur}$ and $\hat{\theta}_r$. The dashed red square is CI_{ur} and the solid blue lines are the boundary of CI_r. In the first panel $\hat{\theta}_{ur} = \hat{\theta}_r = (0, 0)'$. Since $\hat{\theta}_{ur} = \hat{\theta}_r$ and $c_{r,n}(\theta) \leq c_{ur}$ for all $\theta \in \mathbb{R}^2$, it holds that $CI_r \subset CI_{ur}$. Also notice that since $c_{r,n}(\theta)$ depends on θ, CI_r is not a triangle as opposed to the set $CI_{ur} \cap \{\theta \in \mathbb{R} : \theta_1 \leq \theta_2\}$. The second and the third panel display similar situations with $\hat{\theta}_{ur} = \hat{\theta}_r = (0, 0.1)'$ and $\hat{\theta}_{ur} = \hat{\theta}_r = (0, 0.3)'$, respectively. In both cases, $CI_r \subset CI_{ur}$. It also follows from the previous discussion that if $\hat{\theta}_{ur} = \hat{\theta}_r$ and if $\sqrt{n}(\hat{\theta}_{ur,2} - \hat{\theta}_{ur,1})$ is large enough then $CI_{ur} = CI_r$. Consequently, for any fixed θ_0 with $\theta_{0,1} < \theta_{0,2}$, it holds that $P(CI_r = CI_{ur}) \to 1$. However, this equivalence does not hold if θ_0 is at the boundary or close to the boundary. Furthermore, it then holds with positive probability that $CI_{ur} \cap \{\theta \in \mathbb{R} : \theta_1 \leq \theta_2\} = \emptyset$, while CI_r always contains $\hat{\theta}_r$. The fourth panel illustrates that if $\hat{\theta}_{ur} \neq \hat{\theta}_r$, then CI_r is not a subset of CI_{ur}.

The set CI_r is an exact 95% confidence set for θ_0, but it cannot simply be characterized by its extreme points and it can be hard to report with more than two dimensions. Nevertheless, we can use it to construct a rectangular confidence set. To do so, for $j = 1, 2$ define

$$\hat{\theta}^L_{r,j} = \min_{\theta \in CI_r} \theta_j \quad \text{and} \quad \hat{\theta}^U_{r,j} = \max_{\theta \in CI_r} \theta_j$$

and

$$\overline{CI}_r = \left\{ \theta \in \mathbb{R}^2 : \ \theta_1 \leq \theta_2 \text{ and } \hat{\theta}^L_{r,1} \leq \theta_1 \leq \hat{\theta}^U_{r,1} \text{ and } \hat{\theta}^L_{r,2} \leq \theta_2 \leq \hat{\theta}^U_{r,2} \right\}.$$

Then, by construction, $CI_r \subseteq \overline{CI}_r$ and thus $P(\theta_0 \in \overline{CI}_r) \geq 0.95$. Moreover, just as before, if $\hat{\theta}_{ur} = \hat{\theta}_r$, then $\overline{CI}_r \subseteq CI_{ur}$. If for example $\hat{\theta}_{ur} = \hat{\theta}_r = (0,0)'$, then $\hat{\theta}^U_{r,2} = -\hat{\theta}^L_{r,1} = c_{ur}/\sqrt{n}$ but $\hat{\theta}^U_{r,1} = -\hat{\theta}^L_{r,2} < c_{ur}/\sqrt{n}$, which can be seen from the first panel of Figure 2. Hence, relative to the confidence set from the unrestricted estimator, we obtain width gains for the upper end of the first dimension and the lower end of the second dimension. The width gains decrease as $\hat{\theta}_{ur}$ moves away from the boundary into the interior of Θ_R. Moreover, for any $\hat{\theta}_{ur}$ and $\hat{\theta}_r$ and $j = 1, 2$ we get $\hat{\theta}^U_{r,j} - \hat{\theta}^L_{r,j} \leq 2c_{ur}/\sqrt{n}$. Thus, the sides of the square $\{\theta \in \mathbb{R}^2 : \hat{\theta}^L_{r,1} \leq \theta_1 \leq \hat{\theta}^U_{r,1} \text{ and } \hat{\theta}^L_{r,2} \leq \theta_2 \leq \hat{\theta}^U_{r,2}\}$ are never longer than the sides of the square CI_{ur}. Finally, if $\hat{\theta}_{ur}$ is sufficiently in the interior of Θ_R, then $\overline{CI}_r = CI_{ur}$, which is an important feature of our inference method. We get this equivalence in the interior of Θ_R because we invert a test based on a particular type of test statistic, namely $\max\{|Z_1|, |Z_2|\}$. If we started out with a different test statistic, such as $Z_1^2 + Z_2^2$, we would not obtain $\overline{CI}_r = CI_{ur}$ in the interior of Θ_R. We return to this result more generally in Section 1.3.2 and discuss possible alternative ways of constructing confidence regions in Section 1.8.

This method of constructing confidence sets is easy to generalize. As a first step, let Θ_R be a restricted parameter space and let $Q_n(\theta)$ be a population objective function. Suppose that the unrestricted estimator $\hat{\theta}_{ur}$ minimizes $Q_n(\theta)$. Also suppose that $Q_n(\theta)$ is a quadratic function of θ, which holds for example in the NPIV model, and which implies that $\nabla^2 Q_n(\theta)$ does not depend on θ. Then with $\hat{\Omega} = \nabla^2 Q_n(\theta)$ we get

$$Q_n(\theta) = Q_n(\hat{\theta}_{ur}) + \nabla Q_n(\hat{\theta}_{ur})'(\theta - \hat{\theta}_{ur}) + \frac{1}{2}(\theta - \hat{\theta}_{ur})'\hat{\Omega}(\theta - \hat{\theta}_{ur})$$

and since $\nabla Q_n(\hat{\theta}_{ur}) = 0$ it holds that

$$\hat{\theta}_r = \underset{\theta \in \Theta_R}{\arg\min} \; \|\theta - \hat{\theta}_{ur}\|^2_{\hat{\Omega}}.$$

Hence, $\hat{\theta}_r$ is again simply a projection of $\hat{\theta}_{ur}$ onto Θ_R. As before, we can now use a change of variables and characterize the distribution of $\sqrt{n}(\hat{\theta}_r - \theta_0)$ as a projection of $\sqrt{n}(\hat{\theta}_{ur} - \theta_0)$ onto a local parameter space that depends on θ_0 and n. Thus, when testing $H_0 : \theta_0 = \bar{\theta}$ based on a test statistic that depends on $\sqrt{n}(\hat{\theta}_r - \theta_0)$, we can use the projection of the large sample distribution of $\sqrt{n}(\hat{\theta}_{ur} - \theta_0)$ to calculate the critical values.

1.3 General setup

In this section we discuss a general framework and provide conditions for uniformly valid inference. We start with an informal overview of the inference method and provide the formal assumptions and results in Section 1.3.1. In Section 1.3.2 we discuss rectangular confidence regions for general functions of the parameter vector.

Let $\Theta \subseteq \mathbb{R}^{K_n}$ be the parameter space and let $\Theta_R \subseteq \Theta$ be a restricted parameter space. Inferences focuses on $\theta_0 \in \Theta_R$. In an example discussed in Section 1.4.2 we have

$$\theta_0 = \begin{pmatrix} E(Y \mid X = x_1) & \dots & E(Y \mid X = x_{K_n}) \end{pmatrix}',$$

and K_n increases with the sample size. In this case, the confidence regions we obtain are analogous to the ones in the simple example above. For series estimation we take $\theta_0 \in \mathbb{R}^{K_n}$ such that $g_0(x) \approx p_{K_n}(x)'\theta_0$, where g_0 is an unknown function of interest and $p_{K_n}(x)$ is a vector of basis functions. A rectangular confidence region for certain functions of θ_0 can then be interpreted as a uniform confidence band for g_0; see Section 1.4.3 for details. Even though θ_0 and Θ may depend on the sample size, we omit the subscripts for brevity.

As explained in Section 1.2, in many applications we can obtain a restricted estimator as a projection of an unrestricted estimator onto the restricted parameter space. More generally, we assume that there exist $\hat{\theta}_{ur}$ and $\hat{\theta}_r$ such that $\hat{\theta}_r$ is approximately the projection of $\hat{\theta}_{ur}$ onto Θ_R under some norm $\|\cdot\|_{\hat{\Omega}}$ (see Assumption 1.1 below for a formal statement). Moreover, since the rate of convergence may be slower than $1/\sqrt{n}$, let κ_n be a sequence of numbers such that $\kappa_n \to \infty$ as $n \to \infty$. Then

$$\begin{aligned} \hat{\theta}_r &\approx \arg\min_{\theta \in \Theta_R} \|\theta - \hat{\theta}_{ur}\|^2_{\hat{\Omega}} \\ &= \arg\min_{\theta \in \Theta_R} \|\kappa_n(\theta - \theta_0) - \kappa_n(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}}. \end{aligned}$$

Next define

$$\Lambda_n(\theta_0) = \{\lambda \in \mathbb{R}^{K_n} : \lambda = \kappa_n(\theta - \theta_0) \text{ for some } \theta \in \Theta_R\}.$$

Then

$$\kappa_n(\hat{\theta}_r - \theta_0) \approx \arg\min_{\lambda \in \Lambda_n(\theta_0)} \|\lambda - \kappa_n(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}}.$$

We will also assume that $\kappa_n(\hat{\theta}_{ur} - \theta_0)$ is approximately $N(0, \Sigma)$ distributed (see Assumption 1.2 for a formal statement) and that we have a consistent estimator of Σ, denoted by $\hat{\Sigma}$.

Now let $Z \sim N(0, I_{K_n \times K_n})$ be independent of $\hat{\Sigma}$ and $\hat{\Omega}$ and define

$$Z_n(\theta, \hat{\Sigma}, \hat{\Omega}) = \arg\min_{\lambda \in \Lambda_n(\theta)} \|\lambda - \hat{\Sigma}^{1/2} Z\|^2_{\hat{\Omega}}.$$

We will use the distribution of $Z_n(\theta_0, \hat{\Sigma}, \hat{\Omega})$ to approximate the distribution of $\kappa_n(\hat{\theta}_r - \theta_0)$. This idea is similar to Andrews (1999, 2001); see for example Theorem 3 in Andrews (1999). However, our arguments to approximate the large sample distribution are different because we allow the dimension of θ_0 to grow as $n \to \infty$. In addition, we allow θ_0 to be close to the boundary and our local parameter space $\Lambda_n(\theta_0)$, which plays a key role in constructing the critical values, depends on n.[5]

[5] The asymptotic distribution derived in Theorem 3 in Andrews (1999) is a projection of a normal random vector onto a convex cone, which holds quite generally for any fixed θ_0. However, when θ_0 is close to the boundary, such as $(0, 1/\sqrt{n})'$ in the example in Section 1.2, the asymptotic distribution is typically not a projection onto a cone. For example, $\{\lambda \in \mathbb{R}^2 : \lambda_1 - \lambda_2 \leq 1\}$ is not a cone.

Now for $\bar{\theta} \in \Theta_R$ consider testing

$$H_0 : \theta_0 = \bar{\theta}$$

based on a test statistic T, which depends on $\kappa_n(\hat{\theta}_r - \bar{\theta})$ and $\hat{\Sigma}$. For example

$$T(\kappa_n(\hat{\theta}_r - \bar{\theta}), \hat{\Sigma}) = \max_{k=1,\ldots,K_n} \left| \frac{\kappa_n(\hat{\theta}_{r,k} - \bar{\theta}_k)}{\sqrt{\hat{\Sigma}_{kk}}} \right|.$$

We reject H_0 if and only if

$$T(\kappa_n(\hat{\theta}_r - \bar{\theta}), \hat{\Sigma}) > c_{1-\alpha,n}(\bar{\theta}, \hat{\Sigma}, \hat{\Omega}),$$

where

$$c_{1-\alpha,n}(\bar{\theta}, \hat{\Sigma}, \hat{\Omega}) = \inf\{c \in \mathbb{R} : P(T(Z_n(\bar{\theta}, \hat{\Sigma}, \hat{\Omega}), \hat{\Sigma}) \leq c \mid \hat{\Sigma}, \hat{\Omega}) \geq 1 - \alpha\}.$$

Our $1 - \alpha$ confidence set for θ_0 is then

$$CI = \{\theta \in \Theta_R : T(\kappa_n(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega})\}.$$

To guarantee that $P(\theta_0 \in CI) \to 1 - \alpha$ uniformly over a class of distributions $\mathcal{P}$ we require

$$\sup_{P \in \mathcal{P}} \left| P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})\right) - (1 - \alpha) \right| \to 0.$$

Notice that if $\hat{\theta}_r$ was exactly the projection of $\hat{\theta}_{ur}$ onto Θ_R, if $\kappa_n(\hat{\theta}_{ur} - \theta_0)$ was exactly $N(0, \Sigma)$ distributed, if Σ and Ω were known, and if $T(Z_n(\theta_0, \Sigma, \Omega), \Sigma)$ was continuously distributed, then by construction

$$P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \Sigma) \leq c_{1-\alpha,n}(\theta_0, \Sigma, \Omega)\right) = 1 - \alpha,$$

just as in the simple example in Section 1.2. Therefore, the assumptions below simply guarantee that the various approximation errors are small and that small approximation errors only have a small impact on the distribution of the test statistic.

1.3.1 Assumptions and main result

Let ε_n be a sequence of positive numbers with $\varepsilon_n \to 0$. We discuss the role of ε_n after stating the assumptions. Let $\mathcal{P}$ be a set of distributions satisfying the following assumptions.[6]

Assumption 1.1. *There exists a symmetric, positive semi-definite matrix $\hat{\Omega}$ such that*

$$\kappa_n(\hat{\theta}_r - \theta_0) = \underset{\lambda \in \Lambda_n(\theta_0)}{\arg\min} \|\lambda - \kappa_n(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}} + R_n$$

and $\|R_n\| = o_p(\varepsilon_n)$ uniformly over $P \in \mathcal{P}$.

Assumption 1.2. *There exist symmetric, positive definite matrices Ω and Σ and a sequence of random variables $Z_n \sim N(0, \Sigma)$ such that $\lambda_{\min}(\Omega)^{-1/2}\|\kappa_n(\hat{\theta}_{ur} - \theta_0) - Z_n\| = o_p(\varepsilon_n)$ uniformly over $P \in \mathcal{P}$.*

Assumption 1.3. *There exists a constant $C_\lambda > 0$ such that $1/C_\lambda \leq \lambda_{\min}(\Sigma) \leq C_\lambda$, $1/C_\lambda \leq \lambda_{\max}(\Omega) \leq C_\lambda$ for all $P \in \mathcal{P}$ and*

$$\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\hat{\Sigma} - \Sigma\|^2_S = o_p(\varepsilon_n^2/K_n) \qquad \text{and} \qquad \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)^2}\|\hat{\Omega} - \Omega\|_S = o_p(\varepsilon_n^2/K_n)$$

uniformly over $P \in \mathcal{P}$.

Assumption 1.4. *Θ_R is closed and convex and $\theta_0 \in \Theta_R$.*

Assumption 1.5. *Let Σ_1 and Σ_2 be any symmetric and positive definite matrices such that $1/B \leq \lambda_{\min}(\Sigma_1) \leq B$ and $1/B \leq \lambda_{\min}(\Sigma_2) \leq B$ for some constant $B > 0$. There exists a constant C, possibly depending on B, such that for any $z_1 \in \mathbb{R}^{K_n}$ and $z_2 \in \mathbb{R}^{K_n}$*

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

[6] Even though θ_0 depends on $P \in \mathcal{P}$, we do not make the dependence explicit in the notation.

Assumption 1.6. *There exists $\delta \in (0, \alpha)$ such that for all $\beta \in [\alpha - \delta, \alpha + \delta]$*

$$\sup_{P \in \mathcal{P}} \left| P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\beta,n}(\theta_0, \Sigma, \Omega) - \varepsilon_n \right) - (1 - \beta) \right| \to 0$$

and

$$\sup_{P \in \mathcal{P}} \left| P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\beta,n}(\theta_0, \Sigma, \Omega) + \varepsilon_n \right) - (1 - \beta) \right| \to 0.$$

As demonstrated above, if $\hat{\theta}_{ur}$ maximizes $Q_n(\theta)$ and if $\nabla^2 Q_n(\theta)$ does not depend on θ, then Assumption 1.1 holds with $R_n = 0$ and $\hat{\Omega} = \nabla^2 Q_n(\theta)$. Andrews (1999) provides general sufficient conditions for a small remainder in a quadratic expansion. The assumption also holds by construction if we simply project $\hat{\theta}_{ur}$ onto Θ_R to obtain $\hat{\theta}_r$. More generally, the assumption does not necessarily require $\hat{\theta}_{ur}$ to be an unrestricted estimator of a criterion function, which may not even exist in some settings if the criterion function is not defined outside of Θ_R. Even in these cases, $\hat{\theta}_r$ is usually an approximate projection of an asymptotically normally distributed estimator onto Θ_R.[7] Assumption 1.2 can be verified using a coupling argument and the rate of convergence of $\hat{\theta}_{ur}$ can be slower than $1/\sqrt{n}$. Assumption 1.3 ensures that the estimation errors of $\hat{\Sigma}$ and $\hat{\Omega}$ are negligible. If $\lambda_{\min}(\Omega)$ is bounded away from 0 and if $\lambda_{\max}(\Sigma)$ is bounded, then the assumption simply states that $\|\hat{\Sigma} - \Sigma\|_S = o_p(\varepsilon_n/\sqrt{K_n})$ and $\|\hat{\Omega} - \Omega\|_S = o_p(\varepsilon_n^2/K_n)$, which is easy to verify in specific examples. Allowing $\lambda_{\min}(\Omega) \to 0$ is important for ill-posed inverse problems such as NPIV. We explain in Sections 1.4 and 1.5 that both $1/C_\lambda \leq \lambda_{\min}(\Sigma) \leq C_\lambda$ and $1/C_\lambda \leq \lambda_{\max}(\Omega) \leq C_\lambda$ hold under common assumptions in a variety of settings. We could adapt the assumptions to allow for $\lambda_{\min}(\Sigma) \to 0$ and $\lambda_{\max}(\Omega) \to \infty$, but this would require much more notation. Assumption 1.4 holds for example with linear inequality constraints of the form $\Theta_R = \{\theta \in \mathbb{R}^{K_n} : A\theta \leq b\}$. Other examples of convex shape restrictions for series estimators are monotonicity, convexity/concavity, increasing returns to scale, subadditivity,

[7] See Ketz (2018) for the construction of such an estimator. $\hat{\theta}_{ur}$ does not even have to be a feasible estimator and we could simply replace $\kappa_n(\hat{\theta}_{ur} - \theta_0)$ by a random variable $\hat{Z}$, which is allowed for by our general formulation; specifically see Z_T in Andrews (1999).

or homogeneity of a certain degree, but we rule out Slutzki restrictions, which Horowitz and Lee (2017) allow for. The assumption implies that $\Lambda_n(\theta_0)$ is closed and convex as well. The main purpose of this assumption is to ensure that the projection onto $\Lambda_n(\theta_0)$ is nonexpansive, and thus, we could replace it with a higher level assumption, which might then also allow for the Slutzki restrictions.[8] Assumption 1.5 imposes continuity conditions on the test statistic. We provide several examples of test statistics satisfying this assumption in Sections 1.4 and 1.5. Assumption 1.6 is a continuity condition on the distribution of $T(Z_n(\theta_0, \Sigma, \Omega), \Sigma)$, which requires that its distribution function does not become too steep too quickly as n increases. It is usually referred to as an anti-concentration condition and it is not uncommon in these type of testing problems; see e.g. Assumption 6.7 of Chernozhukov et al. (2015). If the distribution function is continuous for any fixed K_n, then the assumption is an abstract rate condition on how fast K_n can diverge relative to ε_n. As explained below, to get around this assumption we could take $c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n$ instead of $c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega})$ as the critical value. Also notice that Assumptions 1.1 – 1.5 impose very little restrictions on the shape restrictions and hence, they are insufficient to guarantee that the distribution function of $T(Z_n(\theta_0, \Sigma, \Omega), \Sigma)$ is continuous.

We now get the following result.

Theorem 1.1. *Suppose Assumptions 1.1 – 1.5 hold. Then*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \geq 1 - \alpha.$$

If in addition Assumption 1.6 holds then

$$\sup_{P\in\mathcal{P}} \left| P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})\right) - (1-\alpha)\right| \to 0.$$

The first part of Theorem 1.1 implies that if we take $c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega}) + \varepsilon$ for any fixed $\varepsilon > 0$ as the critical value, then the rejection probability is asymptotically at most α under the null

[8] I.e. we use $\| \arg\min_{\lambda\in\Lambda_n(\theta_0)} \|\lambda - z_1\|_{\hat{\Omega}} - \arg\min_{\lambda\in\Lambda_n(\theta_0)} \|\lambda - z_2\|_{\hat{\Omega}}\|_{\hat{\Omega}} \leq C\|z_1 - z_2\|_{\hat{\Omega}}$ for some $C > 0$.

hypothesis, even if Assumption 1.6 does not hold. In this case, ε_n can go to 0 arbitrarily slowly. An alternative interpretation is that with $c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega})$ as the critical value and without Assumption 1.6, the rejection probability might be larger than α in the limit, but the resulting confidence set is arbitrarily close to the $1-\alpha$ confidence set. The second part states that the test has the right size asymptotically if Assumptions 1.1 – 1.6 hold.

1.3.2 Rectangular confidence sets for functions

The previous results yield asymptotically valid confidence regions for θ_0. However, these regions might be hard to report if K_n is large and they may not be the main object of interest. For example, we might be more interested in a uniform confidence band for a function rather than a confidence region of the coefficients in the series expansion. We now discuss how we can use these regions to obtain rectangular confidence sets for functions $h : \mathbb{R}^{K_n} \to \mathbb{R}^{L_n}$ using projections, similar as in Section 1.2 where we used $h(\theta) = \theta$. Rectangular confidence regions are easy to report because we only have to report the extreme points of each coordinate, which is crucial when L_n is large.[9] This setup covers both confidence intervals for functionals, such as function values, average derivatives, or an element of θ_0, as well as uniform confidence bands which we focus on in our applications (see Sections 1.4 and 1.5 for the specific function h used in this case). Now define

$$CI = \{\theta \in \Theta_R : T(\kappa_n(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega})\}$$

and let

$$\hat{h}_l^L = \inf_{\theta \in CI} h_l(\theta) \quad \text{and} \quad \hat{h}_l^U = \sup_{\theta \in CI} h_l(\theta), \qquad l = 1, \ldots, L_n.$$

[9]We do not impose any restrictions on L_n and in theory we could have $L_n = \infty$. For example, uniform confidence bands are projections of functionals of the form $p_{K_n}(x)'\theta$ for possibly infinitely many values of x. However, in practice, L_n is typically finite. For example, we can calculate the uniform confidence bands on an arbitrarily fine grid. See Section Sections 1.4 and 1.5 for details.

Notice that if $\theta_0 \in CI$, then $\hat{h}_l^L \leq h_l(\theta_0)$ and $\hat{h}_l^U \geq h_l(\theta_0)$ for all $l = 1, \ldots, L_n$. We therefore obtain the following corollary.[10]

Corollary 1.1. *Suppose Assumptions 1.1 – 1.6 hold. Then*

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(\hat{h}_l^L \leq h_l(\theta_0) \leq \hat{h}_l^U \text{ for all } l = 1, \ldots, L_n\right) \geq 1 - \alpha.$$

A projection for any T satisfying the assumptions above yields a rectangular confidence region with coverage probability at least $1-\alpha$ in the limit. In the examples discussed in Sections 1.4 and 1.5 we pick T such that the resulting confidence region is nonconservative for θ_0 in the interior of Θ_R, just as the confidence sets in Figure 2. In these examples $h_l(\theta) = c_l + q_l'\theta$, where c_l is a constant and $q_l \in \mathbb{R}^{L_n}$, and possibly $L_n > K_n$. We then let

$$T(\kappa_n(\hat{\theta}_r - \theta), \hat{\Sigma}) = \sup_{l=1,\ldots,L_n} \left\{\kappa_n \left|q_l'(\hat{\theta}_r - \theta)\right| / \sqrt{q_l'\hat{\Sigma}q_l}\right\}.$$

Now suppose that for any $\theta \in CI$, the critical value does not depend on θ, which will be the case with probability approaching 1 if θ_0 is in the interior of the parameter space. That is $c(\theta, \hat{\Sigma}, \hat{\Omega}) = \hat{c}$. Then

$$CI = \left\{\theta \in \Theta_R : h_l(\hat{\theta}_r) - \frac{\hat{c}}{\kappa_n}\sqrt{q_l'\hat{\Sigma}q_l} \leq h_l(\theta) \leq h_l(\hat{\theta}_r) + \frac{\hat{c}}{\kappa_n}\sqrt{q_l'\hat{\Sigma}q_l}\ \forall l = 1, \ldots, L_n\right\}.$$

Moreover, by the definitions of the infimum and the supremum as the largest lower bound and smallest upper bound respectively, it holds that

$$\hat{h}_l^L \geq h_l(\hat{\theta}_r) - \frac{\hat{c}}{\kappa_n}\sqrt{q_l'\hat{\Sigma}q_l} \quad \text{and} \quad \hat{h}_l^U \leq h_l(\hat{\theta}_r) + \frac{\hat{c}}{\kappa_n}\sqrt{q_l'\hat{\Sigma}q_l}$$

for all $l = 1, \ldots, L_n$ and thus,

$$\hat{h}_l^L \leq h_l(\theta_0) \leq \hat{h}_l^U \text{ for all } l = 1, \ldots, L_n \quad \Longleftrightarrow \quad \theta_0 \in CI.$$

[10] Under Assumptions 1.1 - 1.5 only, we could project onto $\{\theta \in \Theta_R : T(\kappa_n(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\}$ to obtain the same conclusion as in Corollary 1.1.

Consequently

$$P\left(\hat{h}_l^L \leq h_l(\theta_0) \leq \hat{h}_l^U \text{ for all } l = 1, \ldots, L_n\right) = P\left(\theta_0 \in CI\right).$$

In Corollary A.1 in the appendix we state a formal result, which guarantees that the projection based confidence set does not suffer from over-coverage if θ_0 is sufficiently in the interior of the parameter space. The results can be extended to nonlinear functions h along the lines of Freyberger and Rai (2018). We discuss computational details in Section 1.6.

1.4 Conditional mean estimation

In this section we provide sufficient conditions for Assumptions 1.1 – 1.5 when

$$Y = g_0(X) + U, \quad E(U \mid X) = 0$$

and Y, X and U are scalar random variables. We also explain how we can use the projection results to obtain uniform confidence bands for g_0. We first assume that X is discretely distributed to illustrate that the inference method can easily be applied to finite dimensional models. We then let X be continuously distributed and discuss both kernel and series estimators. Throughout, we assume that the data is a random sample $\{Y_i, X_i\}_{i=1}^n$. The proofs of all results in this and the following section are in the supplementary appendix.

1.4.1 Discrete regressors

Suppose that X is discretely distributed with support $\mathcal{X} = \{x_1, \ldots, x_K\}$, where K is fixed. Let

$$\theta_0 = \begin{pmatrix} E(Y \mid X = x_1) & \ldots & E(Y \mid X = x_K) \end{pmatrix}'$$

and

$$\hat{\theta}_{ur} = \begin{pmatrix} \frac{\sum_{i=1}^n Y_i \mathbf{1}(X_i = x_1)}{\sum_{i=1}^n \mathbf{1}(X_i = x_1)} & \ldots & \frac{\sum_{i=1}^n Y_i \mathbf{1}(X_i = x_K)}{\sum_{i=1}^n \mathbf{1}(X_i = x_K)} \end{pmatrix}'.$$

Define $\sigma^2(x_k) = Var(U \mid X = x_k)$ and $p(x_k) = P(X = x_k) > 0$, and let

$$\Sigma = diag\left(\frac{\sigma^2(x_1)}{p(x_1)}, \ldots, \frac{\sigma^2(x_K)}{p(x_K)}\right) \quad \text{and} \quad \hat{\Sigma} = diag\left(\frac{\hat{\sigma}^2(x_1)}{\hat{p}(x_1)}, \ldots, \frac{\hat{\sigma}^2(x_K)}{\hat{p}(x_K)}\right),$$

where $\hat{p}(x_k) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i = x_k)$ and

$$\hat{\sigma}^2(x_k) = \frac{\sum_{i=1}^n Y_i^2 \mathbf{1}(X_i = x_k)}{\sum_{i=1}^n \mathbf{1}(X_i = x_k)} - \left(\frac{\sum_{i=1}^n Y_i \mathbf{1}(X_i = x_k)}{\sum_{i=1}^n \mathbf{1}(X_i = x_k)}\right)^2.$$

Let Θ_R be a convex subset of $\mathbb{R}^K$, such as $\Theta_R = \{\theta \in \mathbb{R}^K : A\theta \leq b\}$. Now define

$$\hat{\theta}_r = \underset{\theta \in \Theta_R}{\arg\min} \|\theta - \hat{\theta}_{ur}\|^2_{\hat{\Sigma}^{-1}}$$

and hence $\hat{\Omega} = \hat{\Sigma}^{-1}$. Other weight functions $\hat{\Omega}$, such as the identity matrix, are possible choices as well. We discuss this issue further in Section 1.8. As a test statistic we use

$$T(z, \hat{\Sigma}) = \max\left\{|z_1|/\sqrt{\hat{\Sigma}_{11}}, \ldots, |z_K|/\sqrt{\hat{\Sigma}_{KK}}\right\}$$

because the resulting confidence region of the unrestricted estimator is rectangular, analogous to the one in Section 1.2. We now get the following result.

Theorem 1.2. *Let $\mathcal{P}$ be the class of distributions satisfying the following assumptions.*

1. *$\{Y_i, X_i\}_{i=1}^n$ is an iid sample from the distribution of (Y, X) with $\sigma^2(x_k) \in [1/C, C]$, $p(x_k) \geq 1/C$, and $E(U^4 \mid X = x_k) \leq C$ for all $k = 1, \ldots, K$ and for some $C > 0$.*

2. *Θ_R is closed and convex and $\theta_0 \in \Theta_R$.*

3. *$\frac{1}{\sqrt{n}} = o(\varepsilon_n^3)$.*

Then Assumptions 1.1 – 1.5 hold and

$$\liminf_{n\to\infty} \inf_{P \in \mathcal{P}} P\left(T(\sqrt{n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \geq 1 - \alpha.$$

If in addition Assumption 1.6 holds then

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

Next let $h_l(\theta) = \theta_l$ for $l = 1, \ldots, K$. Then the results in Section 1.3.2 yield a rectangular confidence region for θ_0, which can be interpreted as a uniform confidence band for $g_0(x_1), \ldots, g_0(x_K)$. Moreover, Corollary A.1 in the appendix shows that the band is nonconservative if θ_0 is sufficiently in the interior of the parameter space.

1.4.2 Kernel regression

We now suppose that X is continuously distributed with density f_X. We denote its support by $\mathcal{X}$ and assume that $\mathcal{X} = [\underline{x}, \overline{x}]$. Let $\{x_1, \ldots, x_{K_n}\} \subset \mathcal{X}$ and

$$\theta_0 = \begin{pmatrix} E(Y \mid X = x_1) & \ldots & E(Y \mid X = x_{K_n}) \end{pmatrix}'.$$

Here K_n increases as the sample size increases and thus, our setup is very similar to Horowitz and Lee (2017). Let $K(\cdot)$ be a kernel function and h_n the bandwidth. The unrestricted estimator is

$$\hat{\theta}_{ur} = \begin{pmatrix} \frac{\sum_{i=1}^n Y_i K\left(\frac{x_1 - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_1 - X_i}{h_n}\right)} & \ldots & \frac{\sum_{i=1}^n Y_i K\left(\frac{x_{K_n} - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_{K_n} - X_i}{h_n}\right)} \end{pmatrix}'.$$

Define $B = \int_{-1}^1 K(u)^2 du$ and $\sigma^2(x) = Var(U \mid X = x)$ and let

$$\Sigma = diag\left(\frac{\sigma^2(x_1)B}{f_X(x_1)}, \ldots, \frac{\sigma^2(x_{K_n})B}{f_X(x_{K_n})}\right) \quad \text{and} \quad \hat{\Sigma} = diag\left(\frac{\hat{\sigma}^2(x_1)B}{\hat{f}_X(x_1)}, \ldots, \frac{\hat{\sigma}^2(x_{K_n})B}{\hat{f}_X(x_{K_n})}\right),$$

where $\hat{f}_X(x_k) = \frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x_k - X_i}{h_n}\right)$ and

$$\hat{\sigma}^2(x_k) = \frac{\sum_{i=1}^n Y_i^2 K\left(\frac{x_k - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_k - X_i}{h_n}\right)} - \left(\frac{\sum_{i=1}^n Y_i K\left(\frac{x_k - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_k - X_i}{h_n}\right)}\right)^2.$$

Just as before, let Θ_R be convex such as $\Theta_R = \{\theta \in \mathbb{R}^{K_n} : A\theta \leq b\}$ and define

$$\hat{\theta}_r = \arg\min_{\theta \in \Theta_R} \|\theta - \hat{\theta}_{ur}\|^2_{\hat{\Sigma}^{-1}},$$

implying that $\hat{\Omega} = \hat{\Sigma}^{-1}$. Finally, as before we let

$$T(z, \hat{\Sigma}) = \max\left\{|z_1|/\sqrt{\hat{\Sigma}_{11}}, \ldots, |z_{K_n}|/\sqrt{\hat{\Sigma}_{K_n K_n}}\right\}.$$

We get the following result.

Theorem 1.3. *Let $\mathcal{P}$ be the class of distributions satisfying the following assumptions.*

1. *The data $\{Y_i, X_i\}_{i=1}^n$ is an iid sample where $\mathcal{X} = [\underline{x}, \overline{x}]$.*

    (a) *$g_0(x)$ and $f_X(x)$ are twice continuously differentiable with uniformly bounded function values and derivatives. $\inf_{x\in\mathcal{X}} f_X(x) \geq 1/C$ for some $C > 0$.*

    (b) *$\sigma^2(x)$ is twice continuously differentiable, the function and derivatives are uniformly bounded on $\mathcal{X}$, and $\inf_{x\in\mathcal{X}} \sigma^2(x) \geq 1/C$ for some $C > 0$.*

    (c) *$E(Y^4 \mid X = x) \leq C$ for some $C > 0$.*

2. *$x_k - x_{k-1} > 2h_n$ for all k and $x_1 > \underline{x} + h_n$ and $x_{K_n} < \overline{x} - h_n$.*

3. *$K(\cdot)$ is a bounded and symmetric pdf with support $[-1, 1]$.*

4. *Θ_R is closed and convex and $\theta_0 \in \Theta_R$.*

5. *$K_n h_n^5 n = o(\varepsilon_n^2)$ and $\frac{K_n^{5/2}}{\sqrt{nh_n}} = o(\varepsilon_n^3)$.*

Then Assumptions 1.1 – 1.5 hold and

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(T\left(\sqrt{nh_n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}\right) \leq c_{1-\alpha,n}\left(\theta_0, \hat{\Sigma}, \hat{\Omega}\right) + \varepsilon_n\right) \geq 1 - \alpha.$$

If in addition Assumption 1.6 holds then

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

The first assumption contains standard smoothness and moment conditions. The second assumption guarantees that estimators of $g_0(x_k)$ and $g_0(x_l)$ for $k \neq l$ are independent, just as in Horowitz and Lee (2017), and it also avoids complications associated with x_k being too close to the boundary of the support. The third assumption imposes standard restrictions on the kernel function and the fourth assumption has been discussed before. The fifth assumption contains rate conditions. Notice that with a fixed K_n, these rates are the standard conditions for asymptotic normality with undersmoothing in kernel regression. The rate conditions also imply that $K_n h_n \to 0$, which is similar to Horowitz and Lee (2017).

Once again with $h_l(\theta) = \theta_l$ for $l = 1, \ldots, K_n$ the results in Section 1.3.2 yield a rectangular confidence region for θ_0, which is a uniform confidence band for $g_0(x_1), \ldots, g_0(x_{K_n})$. Just as in Horowitz and Lee (2017), the band only covers the function at these grid points, but the grid becomes finer as the sample size increases. In the next section, we discuss a series estimator where the confidence bands cover the entire function.

Remark 1.1. *While we use the Nadaraya-Watson estimator for simplicity, the general theory also applies to other estimators, such as local polynomial estimators. Another possibility is to use a bias corrected estimator and the adjusted standard errors suggested by Calonico et al. (2017). Finally, the general theory can also be adapted to incorporate a worst-case bias as in Armstrong and Kolesár (2016) instead of using the undersmoothing assumption; see Section A.3 for details.*

1.4.3 Series regression

In this section we again assume that $X \in \mathcal{X}$ is continuously distributed, but we use a series estimator. One advantage of a series estimator is that it yields uniform confidence bands for the entire function g_0, rather than just a vector of function values.

Let $p_{K_n}(x) \in \mathbb{R}^{K_n}$ be a vector of basis functions and write $g_0(x) \approx p_{K_n}(x)'\theta_0$ for some $\theta_0 \in \Theta_R$. We again let Θ_R be a convex set such as $\{\theta \in \mathbb{R}^{K_n} : A\theta \leq b\}$. For example, we could impose the constraints $\nabla p_{K_n}(x_j)'\theta \geq 0$ for $j = 1, \ldots, J_n$. Notice that J_n is not restricted, and we could even impose $\nabla p_{K_n}(x)'\theta \geq 0$ for all $x \in \mathcal{X}$ if it is computationally feasible.[11] The

[11]For example, with quadratic splines $\nabla p_{K_n}(x)'\theta \geq 0$ reduces to finitely many inequality constraints.

unrestricted and restricted estimators are

$$\hat{\theta}_{ur} = \underset{\theta \in \mathbb{R}^{K_n}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - p_{K_n}(X_i)'\theta\right)^2$$

and

$$\hat{\theta}_{r} = \underset{\theta \in \Theta_R}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - p_{K_n}(X_i)'\theta\right)^2,$$

respectively. The assumptions ensure that both minimizers are unique with probability approaching 1. Since the objective function is quadratic in θ_0 we have

$$\sqrt{n}(\hat{\theta}_r - \theta_0) = \underset{\lambda \in \Lambda_n(\theta_0)}{\arg\min} \|\lambda - \sqrt{n}(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}},$$

where $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} p_{K_n}(X_i)p_{K_n}(X_i)'$ and $\Omega = E(\hat{\Omega})$. Define

$$\Sigma = \left(E(p_{K_n}(X_i)p_{K_n}(X_i)')\right)^{-1} E(U_i^2 p_{K_n}(X_i)p_{K_n}(X_i)') \left(E(p_{K_n}(X_i)p_{K_n}(X_i)')\right)^{-1}.$$

Also let $\hat{U}_i = Y_i - p_{K_n}(X_i)'\hat{\theta}_{ur}$ and

$$\hat{\Sigma} = \hat{\Omega}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2 p_{K_n}(X_i)p_{K_n}(X_i)'\right)\hat{\Omega}^{-1}.$$

Let $\hat{\sigma}(x) = \sqrt{p_{K_n}(x)'\hat{\Sigma}p_{K_n}(x)}$. We use the test statistic

$$T(\sqrt{n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}) = \sup_{x \in \mathcal{X}} \left| \frac{p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r - \theta_0)\right)}{\hat{\sigma}(x)} \right|.$$

The following theorem provides conditions to ensure that confidence sets for θ_0 have the correct coverage asymptotically. We then explain how we can use these sets to construct uniform confidence bands for $g_0(x)$. To state the theorem, let $\xi(K_n) = \sup_{x \in \mathcal{X}} \|p_{K_n}(x)\|$.

Theorem 1.4. *Let $\mathcal{P}$ be the class of distributions satisfying the following assumptions.*

1. *The data $\{Y_i, X_i\}_{i=1}^n$ is an iid sample from the distribution of (Y, X) with $E(U^2 \mid X) \in [1/C, C]$ and $E(U^4 \mid X) \leq C$ for some $C > 0$.*

2. *The basis functions* $p_k(\cdot)$ *are orthonormal on* $\mathcal{X}$ *with respect to the* L^2 *norm and* $f_X(x) \in [1/C, C]$ *for all* $x \in \mathcal{X}$ *and some* $C > 0$.

3. Θ_R *is closed and convex and* $\theta_0 \in \Theta_R$ *is such that for some constants* C_g *and* $\gamma > 0$

$$\sup_{x \in X} |g_0(x) - p_{K_n}(x)'\theta_0| \leq C_g K_n^{-\gamma}.$$

4. $nK_n^{-2\gamma} = o(\varepsilon_n^2)$, $\frac{\xi(K_n)^2 K_n^4}{n} = o(\varepsilon_n^6)$, *and* $\frac{\xi(K_n)^4 K_n^3}{n} = o(\varepsilon_n^2)$.

Then Assumptions 1.1 – 1.5 hold and

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(T(\sqrt{n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \geq 1 - \alpha.$$

If in addition Assumption 1.6 holds then

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

The first assumption imposes standard moment conditions. The main role of the second assumption is to guarantee that the minimum eigenvalues of Σ and Ω are bounded and bounded away from 0. The third assumption says that g_0 can be well approximated by a function satisfying the constraints, and the fourth assumption provides rate conditions. For asymptotic normality of nonlinear functionals Newey (1997) assumes that

$$nK_n^{-2\gamma} + \frac{\xi(K_n)^4 K_n^2}{n} \to 0.$$

For orthonormal polynomials $\xi(K_n) = C_p K_n$ and for splines $\xi(K_n) = C_s\sqrt{K_n}$. Thus, our rate conditions are slightly stronger than the ones in Newey (1997), but we also obtain confidence sets for the K_n dimensional vector θ_0, which we can transform to uniform confidence bands for g_0. The last rate condition, $\frac{\xi(K_n)^4 K_n^3}{n} = o(\varepsilon_n^2)$, is not needed under the additional assumption that $var(U_i \mid X_i) = \sigma^2 > 0$.

Remark 1.2. *In a finite dimensional regression framework with* $K_n = K$*, the third assumption always holds and the fourth assumption only requires that* $n \to \infty$*. In this case the second assumption can be replaced with the condition* $\lambda_{\min}(E(p_K(X)p_K(X)')) \geq 1/C$*.*

To obtain a uniform confidence band for $g_0(X)$, define

$$CI = \{\theta \in \Theta_R : T(\sqrt{n}(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega})\}$$

and let

$$\hat{g}_l(x) = \min_{\theta \in CI} \; p_{K_n}(x)'\theta \quad \text{ and } \quad \hat{g}_u(x) = \max_{\theta \in CI} \; p_{K_n}(x)'\theta.$$

Also notice that $\|p_{K_n}(x)\|^2$ is bounded away from 0 if the basis functions contain the constant function. We get the following result.

Corollary 1.2. *Suppose the assumptions of Theorem 1.4 and Assumption 1.6 hold. Further suppose that* $\inf_{x \in \mathcal{X}} \|p_{K_n}(x)\|^2 > 1/C$ *for some constant* $C > 0$*. Then*

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(\hat{g}_l(x) \leq g_0(x) \leq \hat{g}_u(x) \;\; \forall x \in \mathcal{X}\right) \geq 1 - \alpha.$$

Remark 1.3. *Without shape restriction, inverting our test statistic results in a uniform confidence band where the width of the band is proportional to the standard deviation of the estimated function for each* x *and the proportionality factor depends on* K_n*. This band can also be obtained as a projecting onto the underlying confidence set for* θ_0*; see Freyberger and Rai (2018) for this equivalence result. If* θ_0 *is sufficiently in the interior of the parameter space, an application of Corollary A.1 in Appendix A shows that the restricted band is equivalent to that band with probability approaching 1. In this case the projection based band is not conservative.*

Remark 1.4. *Similar as before, Assumption 1.6 is not needed if the band is obtained by projecting onto* $\{\theta \in \Theta_R : T(\sqrt{n}(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\}$

Remark 1.5. *The results can be extended to a partially linear model of the form* $Y = g_0(X_1) + X_2'\gamma_0 + U$*. The parameter vector* θ_0 *would then contain both* γ_0 *and the coefficients of the series approximation of* g_0*. If* γ_0 *is unrestricted and if the test statistic does not depend on* γ_0 *(for example if we construct a confidence band for* g_0*), then the critical value does not depend on* γ_0*, which can reduce the computational costs considerably.*

1.5 Instrumental variables estimation

As the final application of the general method we consider the NPIV model

$$Y = g_0(X) + U, \quad E(U \mid Z) = 0,$$

where X and Z are continuously distributed scalar random variables with bounded support. We assume for notational simplicity that X and Z have the same support, $\mathcal{X}$, but this assumption is without loss of generality because X and Z can always be transformed to have support on $[0, 1]$. We assume that $E(U^2 \mid Z) = \sigma^2$ to focus on the complications resulting from the ill-posed inverse problem. Here, the data is a random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$.

As before, let $p_{K_n}(x) \in \mathbb{R}^{K_n}$ be a vector of basis functions and write $g_0(x) \approx p_{K_n}(x)'\theta_0$ for some $\theta_0 \in \Theta_R$, where Θ_R is a convex subset of $\mathbb{R}^{K_n}$. Let P_X be the $n \times K_n$ matrix, where the ith row is $p_{K_n}(X_i)'$ and define P_Z analogously. Let Y be the $n \times 1$ vector containing Y_i. Let

$$\hat{\theta}_{ur} = \underset{\theta \in \mathbb{R}^{K_n}}{\arg\min}\, (Y - P_X\theta)'P_Z(P_Z'P_Z)^{-1}P_Z'(Y - P_X\theta)$$

and

$$\hat{\theta}_r = \underset{\theta \in \Theta_R}{\arg\min}\, (Y - P_X\theta)'P_Z(P_Z'P_Z)^{-1}P_Z'(Y - P_X\theta).$$

For simplicity we use the same basis function as well as the same number of basis functions for X_i and Z_i. Our results can be generalized to allow for different basis functions and more instruments than regressors. Since the objective function is quadratic in θ_0 we have

$$\sqrt{n}(\hat{\theta}_r - \theta_0) = \underset{\lambda \in \Lambda_n(\theta_0)}{\arg\min} \|\lambda - \sqrt{n}(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}},$$

where $\hat{\Omega} = \frac{1}{n}(P_X' P_Z)(P_Z' P_Z)^{-1}(P_Z' P_X)$. Furthermore, let $Q_{XZ} = E(p_{K_n}(X_i) p_{K_n}(Z_i)')$. Then

$$\Sigma = \sigma^2 Q_{XZ}^{-1} E(p_{K_n}(Z_i) p_{K_n}(Z_i)')(Q_{XZ}')^{-1},$$

which we estimate by $\hat{\Sigma} = \hat{\sigma}^2 \hat{\Omega}^{-1}$ with $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2$ and $\hat{U}_i = Y_i - p_{K_n}(X_i)'\hat{\theta}_{ur}$.

As before, $\hat{\sigma}(x) = \sqrt{p_{K_n}(x)'\hat{\Sigma} p_{K_n}(x)}$ and the test statistic is

$$T(\sqrt{n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}) = \sup_{x \in \mathcal{X}} \left| \frac{p_{K_n}(x)' \left(\sqrt{n}(\hat{\theta}_r - \theta_0)\right)}{\hat{\sigma}(x)} \right|.$$

The following theorem provides conditions to ensure that confidence sets for θ_0 have the correct coverage. Analogously to before we can transform these sets to uniform confidence bands for $g_0(x)$. As before, let $\xi(K_n) = \sup_{x \in \mathcal{X}} \|p_{K_n}(x)\|$.

Theorem 1.5. *Let $\mathcal{P}$ be the class of distributions satisfying the following assumptions.*

1. *The data $\{Y_i, X_i, Z_i\}_{i=1}^n$ is an iid sample from the distribution of (Y, X, Z) with $E(U^2 \mid Z) = \sigma^2 \in [1/C, C]$ and $E(U^4 \mid Z) \leq C$ for some $C > 0$.*

2. *The functions $p_k(\cdot)$ are orthonormal on $\mathcal{X}$ with respect to the L^2 norm and the densities of X and Z are uniformly bounded above and bounded away from 0.*

3. *Θ_R is closed and convex and for some function $b(K_n)$ and $\theta_0 \in \Theta_R$*

$$\sup_{x \in \mathcal{X}} |g_0(x) - p_{K_n}(x)'\theta_0| \leq b(K_n).$$

4. *$\lambda_{\min}(Q_{XZ} Q_{XZ}') \geq \tau_{K_n} > 0$ and $\lambda_{\max}(Q_{XZ} Q_{XZ}') \in [1/C, C]$ for some $C < \infty$.*

5. *$\frac{n b(K_n)^2}{\tau_{K_n}^2} = o(\varepsilon_n^2)$ and $\frac{\xi(K_n)^2 K_n^4}{n \tau_{K_n}^6} = o(\varepsilon_n^6)$.*

Then Assumptions 1.1 – 1.5 hold and

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\left(T(\sqrt{n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \geq 1-\alpha.$$

If in addition Assumption 1.6 holds then

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

Assumptions 1 – 3 of the theorem are very similar to those of Theorem 1.4. Assumption 4 defines a measure of ill-posedness τ_{K_n}. For a fixed distribution of the data, we would have $\tau_{K_n} = \lambda_{\min}(Q_{XZ}Q'_{XZ})$, but here the measure is defined as a lower bound for all distributions in $\mathcal{P}$. It is similar to the measure of ill-posedness in Chen and Christensen (2018) and it plays a similar role because it affects the rate of convergence and the rate conditions in Theorem 1.5. It is easy to show that $\lambda_{\max}(Q_{XZ}Q'_{XZ})$ is bounded as long as f_{XZ} is square integrable. However, $\lambda_{\max}(Q_{XZ}Q'_{XZ}) \leq C$ also allows for $X = Z$ as a special case. In fact, in this case, τ_{K_n} is bounded away from 0 and all assumptions reduce to the ones in the series regression framework with homoskedasticity. Moreover, similar to Remark 1.2, the assumptions also allow for K_n to be fixed in which case all conditions reduce to standard assumptions in a parametric IV framework. Finally, the results can also be extended to a partially linear model; see Remark 1.5.

1.6 Monte Carlo simulations

To investigate finite sample properties we simulate data from the model

$$Y = g_0(X) + U, \quad E(U \mid Z) = 0,$$

where $X \in [-1, 1]$ and

$$g_0(X) = -\frac{c}{\sqrt{n}} F^{-1}\left(\frac{1}{4}X + \frac{1}{2}\right).$$

Here, F is the cdf of a t-distribution with one degree of freedom and we vary the constant c. Figure 3 in Appendix (A.1) shows the function for $n = 5,000$ and $c \in \{0, 10, 20, 30, 40, 50\}$. Clearly, $c = 0$ belongs to the constant function. As c increases the slope of $g_0(x)$ increases for every x.

Let $\tilde{X}$, $\tilde{Z}$, and U be jointly normally distributed with $var(U) = 0.25$ and $var(\tilde{Z}) = var(\tilde{X}) = 1$. Let $X = 2F_{\tilde{X}}(\tilde{X}) - 1 \sim Unif[-1,1]$ and $Z = 2F_{\tilde{Z}}(\tilde{Z}) - 1 \sim Unif[-1,1]$. We consider two DGPs. First, we let $cov(\tilde{X}, U) = 0$. Thus, X is exogenous and we use the series estimator described in Section 1.4.3. Second, we let $cov(\tilde{X}, \tilde{Z}) = 0.7$ and $cov(\tilde{X}, U) = 0.5$ and use the NPIV estimator. In both cases we first focus on uniform confidence bands for g_0. In this section we report results with Legendre polynomials as basis function. In Section A.5 in the supplement we report qualitatively very similar results for quadratic splines. For the series regression setting we take $n = 1,000$ and for NPIV we use $n = 5,000$. We take sample sizes large enough such that the unrestricted estimator has good coverage properties for a sufficiently large number of series terms, which helps in analyzing how conservative the restricted confidence bands can be. All results are based on $1,000$ Monte Carlo simulations.

To obtain the uniform confidence bands, we solve optimization problems of the form

$$\min_{\theta \in \mathbb{R}^{K_n}} p_{K_n}(x_l)'\theta \quad \text{and} \quad \max_{\theta \in \mathbb{R}^{K_n}} p_{K_n}(x_l)'\theta, \qquad l = 1, \ldots, L$$

subject to

$$\sup_{x \in \mathcal{X}} \left| \frac{p_{K_n}(x)' \left(\sqrt{n}(\hat{\theta}_r - \theta) \right)}{\hat{\sigma}(x)} \right| \leq c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega}) \text{ and } \nabla p_{K_n}(x_j)'\theta \leq 0 \text{ for all } j = 1, \ldots, J.$$

The second constraint imposes the restriction that g_0 is weakly decreasing and we enforce it on 10 equally spaced points for the polynomial basis. With quadratic splines the constraint $\nabla p_{K_n}(x)'\theta \leq 0$ for all $x \in [-1, 1]$ reduces to finitely many inequality constraints. We solve for the uniform confidence bands on 30 equally spaced grid point. Using finer grids has almost no impact on the results, but increases the computational costs.[12] To solve the optimization problems, we have to calculate $c_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$, which is not available in closed form. To do so, we take $2,000$ draws from a multivariate normal distribution and use them to estimate the distribution function of $T(Z_n(\theta, \hat{\Sigma}, \hat{\Omega}), \hat{\Sigma})$ using a kernel estimator and Silverman's rule of thumb bandwidth. We then take the $1-\alpha$ quantile of the estimated distribution function as the critical value. Estimating the distribution function simply as a step function yields almost identical critical values for any given θ and thus, a test inversion approach would yield essentially identical confidence sets for θ_0. However, the smooth estimated distribution function performed much more reliably when we solved the optimization problems. The number of draws from the normal distribution is analogous to the number of bootstrap samples in other settings and using more draws has almost no impact on our results. Sections 1.6.1 and A.4 provide additional details and discuss the computational costs.

Tables 1 and 2 in Appendix A.1 show the simulation results for the series regression model and the NPIV model, respectively. The first column is the order of the polynomial and $K_n = 2$ belongs to a linear function. We use the same number of basis functions for X and Z, but using $K_n + 3$ for the instrument matrix yields very similar results. The third and fourth columns show the coverage rates of uniform confidence bands using the unrestricted and shape restricted method, respectively. The nominal coverage rate is 0.95. For a confidence band $[\hat{g}_l(x), \, \hat{g}_u(x)]$ define the average width as $\frac{1}{30}\sum_{j=1}^{30}(\hat{g}_u(x_j) - \hat{g}_l(x_j))$, where $\{x_j\}_{j=1}^{30}$ are the grid points. Columns 5 and 6 show the medians of the average widths of the $1,000$ simulated data sets for the unrestricted and restricted estimator, respectively. Let $width_{ur}^s$ and $width_r^s$ be the average widths in data set s. The last columns shows the median of $(width_{ur}^s - width_r^s)/width_{ur}^s$ across the $1,000$ simulated data sets. Even though the mean gains are very similar, we report

[12]In the application we use a grid of 100 points for the uniform confidence bands, but we use a coarser grid for the simulations, because our reported results are based on $78,000$ estimated confidence bands in total.

the median gains to ensure that our gains are not mainly caused by extreme outcomes.

In Table 1 we can see that the unrestricted estimator has coverage rates close to 0.95 if $c = 0$. For $K_n = 2$ and $K_n = 3$, the coverage probability drops significantly below 0.95 when c is large because increasing c also increases the approximation bias. For larger values of K_n, the coverage probability of the unrestricted band is close to 0.95 for all reported values of c. Due to the projection, the coverage probability of the restricted band tends to be above the one of the unrestricted band. When c is large enough, such as $c = 10$ with $K_n = 2$, the two bands are identical with very large probability. The average width of the unrestricted band does not depend on c. On the other hand, the average width of the restricted band is much smaller when c is small. Consequently, the restricted band can be significantly narrower than the unrestricted band with the same K_n, especially when K_n is large and c is small.[13]

Table 2 shows that the results for the NPIV model are similar, but the gains from using the shape restrictions are much larger. For example, when $K_n = 5$ and $c = 0$, the unrestricted bands tends to be almost four times as wide as the restricted band. Furthermore, the range of c values for which we achieve sizable gains is much larger for in the NPIV model relative to the series regression framework. More generally, due to the larger variance of the estimator in the NPIV model, we observed in a variety of other simulations that the range of functions for which we obtain gains in this model is much larger than in series regression for the same

[13]Since U is normally distributed and independent of X, an alternative method to construct confidence bands in this setting is the one proposed by Dümbgen (2003). Since this method only applies with fixed regressors and since $X \sim Unif(0,1)$ in our simulations, we let $X \in \{-1,000/1,002, -998/1,002, \ldots, 998/1,002\}$, which is an equal spaced grid of size $1,000$. We also assume that the variance of U is known. With the monotonized bands of Dümbgen (2003) we then get coverage rates and widths of 0.948 and 0.233 when $c = 0$, 0.974 and 0.258 when $c = 2$, 0.979 and 0.281 when $c = 4$, 0.983 and 0.301 when $c = 6$, 0.986 and 0.319 when $c = 8$, and 0.988 and 0.336 when $c = 10$. This method is not conservative at the boundary, but it is conservative in the interior. The bands can be empty and arbitrarily small (with and without monotonizing), but they are empty in only 3 out of $1,000$ samples when $c = 0$. An advantage of the method is that it is smoothing parameter free, but the widths are much larger than our widths, even when $K_n = 5$, for all values of c (our method yields almost identical results as those in Table 1 when we fix the grid). Also notice that this method only applies when U is normally distributed and the regressors are fixed.

sample size and a similar DGP.

The widths gains reported in Tables 1 and 2 compare restricted and unrestricted bands for the same values of K_n. Even without shape restrictions, it is not clear how to pick K_n in an optimal, data-driven way, which ensures that the confidence bands have the right coverage probability asymptotically. Investigating a data-driven choice for K_n is therefore out of the scope of this paper, but we should note that the optimal K_n might be different for estimators with and without shape restriction. Hence, an alternative way of interpreting the results in Tables 1 and 2 is to compare the widths of restricted and unrestricted bands with coverage probabilities close to 95%. For example, in the NPIV model and $c = 5$, the coverage probability of the unrestricted estimator is close to 95% when $K_n = 4$ and the median width is then 0.207. For the same DGP, the restricted estimator has coverage probabilities close to or above 95% when $K_n = 3$ and $K_n = 4$ with median widths of 0.079 and 0.120, respectively. These comparisons suggest that the widths gains from the shape restrictions are also large with data-driven choices of K_n when g_0 is close to the boundary.

Figure 4 in Appendix (A.1) shows the means of the restricted and the unrestricted bands obtained from the $1,000$ simulated data sets in the NPIV model with $K_n = 5$. Figure 5 in Appendix (A.1) contains four specific examples when $c = 5$. In the first example, both the restricted and the unrestricted estimator are monotone, but the restricted band is still much smaller. In the last example the unrestricted band does not contain any monotone function. In contrast, the restricted bands are always centered around the restricted estimates and both the upper and lower bound functions are monotonically decreasing.

In addition to some of the information in Table 2 in Appendix (A.1), we report the width gains relative to monotonized bands in Table 3 for a subset of the DGPs with $K_n = 5$. To obtain these bands we simply exclude all parts of the unrestricted bands which are not consistent with a weakly decreasing function. As we can see from the reported widths, our restricted bands are considerably smaller than these bands as well, and the widths gains are up to $1 - 0.122/0.210 = 41.9\%$. Moreover, the monotonized band may be empty, which happens in 1.6% of the data sets when $c = 0$, or they can be extremely narrow.

Finally, to illustrate that our method is also applicable to functionals, Table 4 in Appendix (A.1) shows coverage rates and median widths of confidence intervals for the average derivative of the function. Here, we compare our method to the series estimator without shape restrictions (corresponding to cov_{ur} and $widths_{ur}$ in Table 4), the unrestricted approach, but only including the non-positive part of the confidence interval (corresponding to cov_{ur} and $widths_{neg}$), and the method of Chernozhukov et al. (2015) (corresponding to cov_{cns} and $widths_{cns}$). Again, our method yields considerable width gains compared to the unrestricted intervals at or close to the boundary (up to $1 - 0.233/0.637 = 63\%$ when $c = 0$) and coverage rates above 95%. The approach of Chernozhukov et al. (2015) is not conservative at the boundary, their intervals are in this case narrower than ours (0.164 versus 0.233), and they are empty in 4% of the samples. As we move into the interior of the parameter space, our method performs favorably, which could be due to the choice of the user specified tuning parameters which their approach requires. We use their suggested data dependent procedures and did not explore other choices.[14] The first row of Table 4 contains results when $c = -5$ and thus, the model is misspecified and the true function is increasing. In this case, their approach yields empty intervals in 81% of the cases, while our confidence sets are always built around the restricted estimators.

1.6.1 Computational costs

We ran the simulations using MATLAB and the resources of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. To get accurate

[14] Specifically, we use the "aggressive" data dependent choices for r_n and l_n explained in their Section 7.1. These choices might lead to a choice of r_n which is too large in the interior of the parameter space and thus, confidence intervals that are too conservative and unnecessarily wide.

computational times we rerun a subset of the simulations using MATLAB R2015a on a desktop computer with an Intel Core i3 processor running at 3.5Ghz. In these simulations, the median times to solve for the uniform confidence bands in the NPIV setting were roughly 5 minutes when $K_n = 2$, 15 minutes when $K_n = 3$, 32 minutes when $K_n = 4$, and 45 minutes when $K_n = 5$. Each of these times is based on 60 simulated data sets. In Section A.4 in the supplement, we provide additional details, such as our selection of starting values. Since we use a grid of 30 points to calculate the uniform confidence bands, we solve 60 optimization problems for each band. Therefore, obtaining confidence intervals for the average derivatives is considerably faster. In particular, the median time is around 2.5 minutes, even though $K_n = 5$. The approach of Chernozhukov et al. (2015) is based on test inversion, where we test a particular value for the average derivative and obtain critical values using the bootstrap. With $2,000$ bootstrap samples, which is then comparable to the $2,000$ normal draws we use for our approach, each test takes around 6 seconds. Thus, the computational costs of the two approaches in this particular setting are similar if we use 25 grid points for the test inversion approach of Chernozhukov et al. (2015), which is considerably less than what we used to obtain the results in Table 4.

There are several possibilities to substantially reduce the computational times. First, notice that the programs for the uniform confidence bands are very easy to parallelize because the optimization problems are solved separately for each grid point. Second, in our setting we could also use an approach recently suggested by Kaido et al. (2016) in a computationally similar problem in the moment inequality literature. In our setting, their algorithm leads to essentially identical results in both the simulations and the empirical application; see Section A.4 for more details. Moreover, for the uniform confidence bands in the NPIV setting, the median times with their approach are roughly 2.5 minutes when $K_n = 2$, 8 minutes when $K_n = 3$, 13 minutes when $K_n = 4$, and 20 minutes when $K_n = 5$. Finally, we recently developed the code in Fortran, which runs approximately ten times faster than the MATLAB code in the empirical application below, where we have 16 estimated parameters, and it yields identical results.

1.7 Empirical application

In this section, we use the data from Blundell et al. (2012) and Chetverikov and Wilhelm (2017) to estimate US gasoline demand functions and to provide uniform confidence bands under the assumption that the demand function is weakly decreasing in the price. The data comes from the 2001 National Household Travel Survey and contains, among others, annual gasoline consumption, the gasoline price, and household income for $4,812$ households. We excludes households from Georgia because their gasoline price is much smaller than for all other regions (highest log price of 0.133 while the next largest log price observation is 0.194) and therefore $n = 4,655$.

We use the model

$$Y = g_0(X_1, X_2) + X_3'\gamma_0 + U, \qquad E(U \mid Z, X_2, X_3) = 0.$$

Here Y denotes annual log gasoline consumption of a household, X_1 is the log price of gasoline (the average local price), X_2 is log household income, and X_3 contains additional household characteristics, namely the log age of the household respondent, the log household size, the log number of drivers, and the number of workers in the household. Following Blundell et al. (2012) and Chetverikov and Wilhelm (2017), we use the distance to a major oil platform as an instrument, denoted by Z, for X_1. We report estimates and uniform confidence bands for $g_0(x_1, \bar{x}_2) + \bar{x}_3'\gamma_0$ as a function of x_1. We fix X_3 at their mean values and we consider three different values of $\bar{x}_2$, namely the 25th percentile, the median, and the 75th percentile of the income distribution.

The estimator is similar to the one described in Section 1.5 and our specification is similar to Chetverikov and Wilhelm (2017). Specifically, we use quadratic splines with three interior knots for X_1 (contained in the matrix P_{X_1}) and cubic splines with eight knots for Z (contained in the matrix P_Z). The matrix of regressors is then $(P_{X_1}, P_{X_1} \times X_2, X_3)$, where $\times$ denotes the tensor product of the columns of the matrices, and $(P_Z, P_Z \times X_2, X_3)$ is the matrix of instruments. Chetverikov and Wilhelm (2017) estimate γ in the first step and subtract $X_3'\hat{\gamma}$ from Y, while we estimate all parameters together in order to incorporate the variance of $\hat{\gamma}$ when constructing confidence bands. We also report results for a second specification using quadratic splines with six knots to construct P_Z to illustrate the sensitivity of the estimates.

Figure 6 in Appendix (A.1) plots unrestricted and restricted estimators for the three income levels along with 95% uniform confidence bands. The left side contains the estimates with quadratic splines and six knots for Z and right side the estimates with cubic splines and eight knots. The unrestricted estimates are generally increasing for very low and high prices, suggesting that the true demand function has a relatively small slope for these price levels. Our bands are centered around the restricted estimates with monotone upper and lower bound functions. Moreover, the average width of the restricted band is between 25% and 45% smaller than the average width of the unrestricted band. We can also see from the figures that the unrestricted estimates and bands are very sensitive to the specification, but the restricted ones are not.

1.8 Conclusion

We provide a general approach for conducting uniformly valid inference under shape restrictions. A main application of our method is the estimation of uniform confidence bands for an unknown function of interest, as well as confidence regions for other features of the function.

Our confidence bands are well suited to be reported along with shape restricted estimates, because they are build around restricted estimators and the upper and lower bound functions are consistent with the shape restrictions. In addition, the bands are asymptotically equivalent to standard unrestricted confidence bands if the true function strictly satisfies all shape restrictions, but they can be much smaller if some of the shape restrictions are binding or close to binding. Our method is widely applicable and we provide low level conditions for our assumptions in a regression framework (for both series and kernel estimation) and the NPIV model. We demonstrate in simulations and in an empirical application that our shape restricted confidence bands can be much narrower than unrestricted bands.

There are several interesting directions for future research. First, while we prove uniform size control, we do not provide a formal power analysis. It is known that monotone nonparametric estimators can have a faster rate of convergence if the true function is close to constant (see for example Chetverikov and Wilhelm (2017)). Our simulation results suggest that our bands also converge at a faster rate than unrestricted bands in this case, but establishing this result formally is out of the scope of this paper.

Second, we assume that the restricted estimator is an approximate projection of the unrestricted estimator under a weighted Euclidean norm $\|\cdot\|_{\hat{\Omega}}$. In many settings the matrix $\hat{\Omega}$ can be chosen by the researcher (as in Section 1.4.2). For example, in a just identified GMM setting it is well known that the unrestricted estimator is invariant to the GMM-weight matrix. However, the restricted estimator generally depends on the GMM-weight matrix because it affects $\hat{\Omega}$. Notice that $\hat{\theta}_{ur}$ is approximately $N(\theta_0, \hat{\Sigma}/\kappa_n^2)$ distributed. To obtain the restricted estimator of θ_0 we could image maximizing the likelihood with respect to θ_0, where the data is $\hat{\theta}_{ur}$, subject to the solution being in Θ_R. It is easy to show that the restricted maximum likelihood estimator is $\arg\min_{\theta \in \Theta_R} \|\theta - \hat{\theta}_{ur}\|_{\hat{\Sigma}^{-1}}$, suggesting to use $\hat{\Omega} = \hat{\Sigma}^{-1}$, although it is not clear that MLE has optimality properties in this setting. In a just identified GMM setting, such as our regression or IV framework, this amounts to using the standard optimal GMM-weight matrix. In simulations, we found that this weight matrix performs particularly well, but we leave optimality considerations for future research.

Finally, notice that in our setting $\hat{\theta}_r$ is a function of $\hat{\theta}_{ur}$ and hence, $\hat{\theta}_{ur}$ provides more information than $\hat{\theta}_r$. Therefore, instead of letting the test statistic depend on $\kappa_n(\hat{\theta}_r - \theta_0)$, we could let it depend on $\kappa_n(\hat{\theta}_{ur} - \theta_0)$ and incorporate the shape restrictions in the test statistic. This approach would potentially allow us to use additional test statistics. We are particularly interested in rectangular confidence sets for functions of θ_0, which are equivalent to standard rectangular confidence sets if θ_0 is in the interior of Θ_R. Such sets can be obtained using test statistics that depend on $\kappa_n(\hat{\theta}_r - \theta_0)$ and it is therefore not immediately obvious what the potential benefits of a more general formalization are.

Chapter 2

Simple Inference in First-Price Auctions[1]

2.1 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.

[1] **This chapter is joint work with Cristián Hernández**

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-Raphson step to a restricted minimizer of a sample criterion function. Intuitively, the application of the Newton-Raphson 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 piece-wise pseudo maximum likelihood estimator of Donald and Paarsch (1993). Their method assumes one can partition the parameter θ into a scalar, θ_1, and the remaining parameters, θ_2, where the rightmost support point of the bid distribution is strictly increasing in θ_1. For any value of θ_2, they set θ_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 θ_2 after concentrating out θ_1. Under mild regularity conditions, their estimator for θ_2 is asymptotically normal. Laffont et al. (1995) proposed a second alternative based on simulated method of moments which estimates θ by minimizing a nonlinear least squares objective function measuring the distance between sample moments and simulated moments using θ. 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-Raphson 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 2.2 describes the auction environment and a description of its non-standard features. Section 2.3 introduces our estimator and high-level conditions under which our tests control size. Section 2.4 contains low-level sufficient conditions for a simple auction model. Section 2.5 contains the simulation study. Section 2.6 contains the empirical application. Section 2.7 concludes and discusses avenues for extensions. Additionally, we include an appendix of figures/tables in section B.1 and a technical appendix in section B.2. 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.

2.2 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 2.2.1, we outline the simple auction model we consider, and in subsection 2.2.2 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, 2.3.3 we return to the general environment which allows for both parametric and nonparametric specifications.

2.2.1 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 $[\underline{v}, \overline{v}]$. 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 $f(\cdot, \theta)$. Using standard results in the auctions literature, we can derive the bidding strategy as

$$b(v, \theta, p) = v - \frac{\int_{\underline{v}}^{v} F(t, \theta)^{p-1} dt}{F(v, \theta)^{p-1}}. \tag{2.1}$$

Under the assumption that $f(v, \theta) > 0$ for all $v \in [\underline{v}, \overline{v}]$ 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

$$G(b, \theta, p) = P\left(b(v, \theta, p) \leq b\right) = P\left(v \leq \eta(b, \theta, p)\right) = F(\eta(b, \theta, p), \theta). \tag{2.2}$$

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 (2.2) 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

$$g(b, \theta, p) = \frac{\partial G(b, \theta, p)}{\partial b} = f(\eta(b, \theta, p)) \frac{\partial \eta(b, \theta, p)}{\partial b} = \frac{F(\eta(b, \theta, p), \theta)}{(p-1)(\eta(b, \theta, p) - b)} \tag{2.3}$$

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).

2.2.2 Illustration of the Non-Standard Features

Given expression (2.3), an intuitive estimator of θ 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 2.3.

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 $\overline{v}$, and this agent strategically responds to the entire shape of the valuation distribution. To see this, define

$$\overline{b}(\theta, p) = \lim_{v \to \overline{v}} b(v, \theta, p) = \overline{v} - \int_{\underline{v}}^{\overline{v}} F(t, \theta)^{p-1} dt, \tag{2.4}$$

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 $\{\{b_{il}\}_{i=1}^{p}\}_{l=1}^{L}$ and consider evaluating the log-likelihood for an arbitrary value of θ. By equation (2.4), we see the largest bid the model can rationalize given θ is $\overline{b}(\theta, p)$. If any bid in the sample exceeds $\overline{b}(\theta, p)$, the parameter θ could not have generated the given sample as no agent would rationally bid above $\overline{b}(\theta, p)$ for any valuation in $[\underline{v}, \overline{v}]$ if valuations were distributed as $f(\cdot, \theta)$. Therefore, the sample likelihood at θ is zero, which implies the sample log-likelihood is not well-defined for this θ. If we denote

$$\hat{\Theta}_R \equiv \left\{ \theta \in \Theta \,\middle|\, \overline{b}(\theta, p) \geq \max_{i,l}\{b_{i,l}\} \right\}, \tag{2.5}$$

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 $\overline{b}(\theta_0, p)$. Therefore, the log-likelihood is not well-defined for any parameter θ such that $\overline{b}(\theta, p) < \overline{b}(\theta_0, p)$ with probability approaching one for almost all sample paths.[2] Consequently, if $\overline{b}(\theta, p)$ is continuously differentiable in θ with $\frac{\partial \overline{b}(\theta_0, p)}{\partial \theta} \neq 0$, any open-ball of θ_0 contains points θ for which $\overline{b}(\theta, p) < \overline{b}(\theta_0, p)$.[3] As a result, standard asymptotic arguments which assume θ_0 is an interior point in the parameter space may provide poor approximations to finite sample distribution of the estimator.

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

[3]$F(v, \theta)$ continuously differentiable in θ almost everywhere implies $\overline{b}(\theta, p)$ is continuously differentiable for any p.

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

$$\mathbb{E}_{\theta_0}\left(\frac{\partial \log(g(b,\theta_0,p))}{\partial\theta}\right) = -g(\bar{b}(\theta_0),\theta_0,p)\frac{\partial\bar{b}(\theta_0,p)}{\partial\theta} = \frac{-\frac{\partial\bar{b}(\theta_0,p)}{\partial\theta}}{(p-1)\int_{\underline{v}}^{\overline{v}} F(t,\theta_0)^{p-1}dt} \equiv \mu(\theta_0,p). \tag{2.6}$$

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.

2.3 Proposed Estimator and Testing Procedure

2.3.1 Description of Our Estimator

To address the non-standard features of the log-likelihood in section 2.2, 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

$$Q_n(\theta) = m_n(\theta)' m_n(\theta) \quad \text{where} \quad m_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \log(g(b_i,\theta,p))}{\partial\theta} - \mu(\theta,p)\right)$$

and $\mu(\theta,p)$ is given in equation (2.6) so that $\mathbb{E}_\theta(m_n(\theta)) = 0$. 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

$$\hat{\theta}_n \equiv \hat{\theta}_{\text{mm}} + \left(\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{\partial Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta}\right) \quad \text{where} \quad \hat{\theta}_{\text{mm}} = \arg\min_{\theta\in\hat{\Theta}_R} n\, Q_n(\theta), \tag{2.7}$$

which corresponds to a Newton-Raphson step applied to $\hat{\theta}_{\text{mm}}$. The construction of the estimator only requires expressions for the density and support of the bid distribution as functions of θ, which are available in many auction models. In the application in section 2.6, 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 $\hat{\theta}_{\text{mm}}$ (see equation (2.13) 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 2.3.2, the use of re-centered moments in $Q_n(\theta)$ and the Newton-Raphson step address the two non-standard features of the model outlined in section 2.2.2 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 (2.7) 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 2.3.3.

2.3.2 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 2.3.3. 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

$$\hat{\theta}_{\text{mm}} = \underset{\theta \in \hat{\Theta}_R}{\arg\min}\, n\, Q_n(\theta). \tag{2.8}$$

Due to the presence of the restriction, we cannot analyze the distribution of $\hat{\theta}_{\text{mm}}$ 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 θ_0. Assuming the Taylor-expansion holds exactly (c.f. Assumption 2.1 in section 2.3.3), 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

$$n\, Q_n(\theta) = n\, Q_n(\theta_0) - \frac{1}{2} D_n' J_n^{-1} D_n + \frac{1}{2} \left\| \sqrt{n}\, (\theta - \theta_0) - J_n^{-1} D_n \right\|_{J_n}^2 \tag{2.9}$$

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 (2.9) into (2.8) and recognizing the first two terms in (2.9) do not depend on θ gives

$$\hat{\theta}_{\text{mm}} = \underset{\theta \in \hat{\Theta}_R}{\arg\min} \left\| \sqrt{n}\, (\theta - \theta_0) - J_n^{-1} D_n \right\|_{J_n}^2. \tag{2.10}$$

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

$$D_n = \frac{\partial m_n(\theta_0)}{\partial \theta} \sqrt{n}\, m_n(\theta_0) \xrightarrow{d} N(0, V_D) \quad \text{and} \quad J_n \xrightarrow{p} J \tag{2.11}$$

for symmetric matrices V_D and J (c.f. Assumption 2.2 in section 2.3.3). By performing the change of variables $\lambda = \sqrt{n}\,(\theta - \theta_0)$ in equation (2.10) and assuming non-singularity of J (c.f. Assumption 2.5 in section 2.3.3), we can see that $\sqrt{n}\,(\hat{\theta}_{\text{mm}} - \theta_0)$ is approximately distributed as the projection of the normal random variable $N(0, J^{-1}V_D J^{-1})$ on to the space, $\sqrt{n}\,(\hat{\Theta}_R - \theta_0)$. The use of re-centered moments is crucial for establishing the normality in (2.11) (for further discussion, see section 2.3.4). 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 (2.10).

To introduce the simple modification and show how it restores normality, note that the source of the non-standard features of the distribution of $\hat{\theta}_{\text{mm}}$ is the presence of the restriction in equation (2.10). 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 (2.10) is a quadratic function of θ which is defined for all of Θ. Minimizing this quadratic function over Θ, yields the estimator $\theta_n = \theta_0 + J_n^{-1}\frac{1}{\sqrt{n}}D_n$, which corresponds to a Newton-Raphson step applied to θ_0. Re-arranging this and using the result in equation (2.11) gives

$$\sqrt{n}\,(\theta_n - \theta_0) = J_n^{-1} D_n \xrightarrow{d} N(0, J^{-1}V_D J^{-1}). \tag{2.12}$$

Therefore, the Newton-Raphson step effectively inverts the projection operator in equation (2.10) 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 θ_0 which is unknown. If we minimize the quadratic expansion of $Q_n(\theta)$ which is centered at $\hat{\theta}_{\text{mm}}$ we obtain our estimator which is the plug-in version of the infeasible estimator. Specifically, we get

$$\hat{\theta}_n = \hat{\theta}_{\text{mm}} + \hat{J}_n^{-1}\frac{1}{\sqrt{n}}\hat{D}_n = \hat{\theta}_{\text{mm}} - \left(\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{\partial Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta}\right). \tag{2.13}$$

Although it is not immediately clear from equation (2.13), it is possible to show that under mild continuity conditions on the second derivative of $Q_n(\theta)$ (c.f. Assumption 2.4 in section 2.3.3) the feasible and infeasible estimators share the same asymptotic distribution. Specifically, we have

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

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 $\hat{\theta}_{\text{mm}}$ versus θ_0 is asymptotically irrelevant. In the next section, we provide conditions which formalize this argument.

2.3.3 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.[4] For a function(al) $\psi(\cdot)$, we consider tests of the null-hypothesis $H_0 : \psi(\theta_0) = \overline{\psi}$. 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 θ_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

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

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 ε_n be a sequence of positive numbers converging to zero where we discuss the role of ε_n after stating the assumptions and main result. Let $\mathcal{P}$ denote

[4]We have suppressed the dependence of θ_0 on n for convenience.

the set of probability models over which the following assumptions hold.

Assumption 2.1. ***Quadratic Approximation*** *There exists symmetric matrix J_n and a random variable θ_n satisfying the following assumptions such that*

$$\sqrt{n}\,(\hat{\theta}_{mm} - \theta_0) = \underset{\lambda \in \sqrt{n}\left(\hat{\Theta}_R - \theta_0\right)}{argmin} \left\|\lambda - \sqrt{n}\,(\theta_n - \theta_0)\right\|_{J_n} + r_n$$

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

Assumption 2.2. ***Normality Approximation*** *There exists symmetric matrices Σ_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 2.3. ***Covariance Estimation*** $\|\hat{\Sigma}_n - \Sigma_n\| = o_p(\varepsilon_n K_n^{-\frac{1}{2}})$ *uniformly over $\mathcal{P}$.*

Assumption 2.4. ***Continuity of the Criterion Function*** *For any θ_1 with $\|\theta_1 - \theta_0\|^2 = O_p\left(\frac{n}{K_n}\right)$ uniformly over $\mathcal{P}$,*

$$\left\|\frac{\partial^2 Q_n(\theta_1)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right\| = o_p\left(\varepsilon_n K_n^{-\frac{1}{2}}\right)$$

uniformly over $\mathcal{P}$.

Assumption 2.5. ***Bounded Eigenvalues*** *Uniformly over $\mathcal{P}$ the eigenvalues Σ_n and J_n are bounded and bounded away from zero.*

Assumption 2.6. ***Continuity of the Test Statistic*** *For all symmetric positive definite matrices Σ_1 and Σ_2 (with Σ_1 and Σ_2 comparable) with all eigenvalues in $\left[\frac{1}{B}, B\right]$ for some $B < \infty$ there exists a constant C, possibly depending on B, such that uniformly over $\mathcal{P}$*

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

Assumption 2.7. ***Anti-Concentration*** *There exists a $\delta \in (0, \alpha)$ such that for all $\beta \in [\alpha - \delta, \alpha + \delta]$,*

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

and

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

Theorem 2.1. *Suppose Assumptions (2.1)-(2.7) hold. Then*

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

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 (2.1) 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 (2.2). 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 (2.2) and (2.3)) result in an asymptotically negligible distortion of the rejection probability of the test.

Remark 2: Several comments are in order about the assumptions. Assumptions (2.1) and (2.2) ensure the MM estimator behaves approximately as the projection of a normal random variable on to the sample restriction space. In conjunction with assumption (2.4), these assumptions establish the feasible plug-in estimator $\hat{\theta}_n$ and the infeasible estimator θ_n described in section 2.3.2 share the same asymptotic distribution. Assumption (2.4) is slightly stronger than continuity of the second derivative at θ_0. Assumption (2.6) is a mild continuity condition on the test-statistic which when combined with Assumption (2.5) ensures the approximation errors in Assumptions (2.2) and (2.3) have an asymptotically negligible impact on the distribution of the test-statistic. Assumption (2.5) 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 ε_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 ε_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 (2.7). 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 (2.6), this assumption imposes mild conditions on the rate at which ε_n converges to zero (see, for instance, Chernozhukov et al. (2013) for a discussion).

2.3.4 Motivating the Method of Moments using Re-Centered Moments

In section 2.3.1, 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 2.3.3 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 (2.10) in section 2.3.2 to get

$$\hat{\theta}_{\mathrm{ml}} = \left(\arg\min_{\theta \in \hat{\Theta}_R} \left\| \sqrt{n}\,(\theta - \theta_0) - H_n^{-1} S_n \right\|_{H_n}^2 \right) + r_n \tag{2.14}$$

where

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

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 (2.14) 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 2.2.2 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 (2.2) 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 (2.2) 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 (2.1) and (2.5) from section (2.3.3). 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 (2.5) 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 (2.1)) as the second-order Taylor expansion then corresponds to an indefinite quadratic form which may have no global extrema on $\hat{\Theta}_R$.

2.4 Low-Level Sufficient Conditions for a Simple Auction Model

In this section, we introduce low-level sufficient conditions for the conditions presented in section 2.3.3. We focus on assumptions for the nonparametric environment. In section 2.4.1 we discuss the sequence of sieve spaces we use for estimation, section 2.4.2 describes the data configuration and section 2.4.3 states the low-level conditions.

2.4.1 The Sieve Space

Let $\Theta \subseteq \mathbb{R}^{K_L}$ where L denotes the number of observed auctions. For nonparametric models, we take $\theta_0 \in \mathbb{R}^{K_L}$ such that $f_0(v) \approx f(v, \theta_0)$, where our low-level conditions then impose restrictions on how slowly/quickly K_L is allowed to grow with L.[5] For an arbitrary element $\theta \in \Theta$, the associated density is given by

$$f(v, \theta) = \left(P(v)'\theta\right)^2 \phi(v) \tag{2.15}$$

where $\phi(v)$ is a density function supported on $[\underline{v}, \overline{v}]$ which is specified by the researcher and $P(v)$ is a K_L-dimensional vector of $\phi(v)$-orthonormalized basis functions. For example, $\phi(v)$ could be the uniform distribution on $[\underline{v}, \overline{v}]$ or a truncated parametric distributions (such as the log-normal). When $\phi(v)$ corresponds to a (truncated) parameterized distribution with parameter β, we denote this by $\phi(v, \beta)$.[6] The form of $f(v, \theta)$ is common for sieve density estimators as it three attractive features. First, if the vector $P(v)$ contains the constant function, the sieve then contains the parametric sub-model $\phi(v, \beta)$. Therefore, by setting $\phi(v, \beta)$ to commonly used parametric models such as the log-normal density, researchers can nest these parametric sub-models within a more flexible set of models. Second, $f(v, \theta)$ is always non-negative due to the presence of the squared term. Finally, the use of $\phi(v)-$orthonormal basis functions, $P(v)$, implies $f(v, \theta)$ integrates to one if and only if $\|\theta\| = 1$.

2.4.2 Data Configuration and Criterion Function

The data, $\{\boldsymbol{b}_l\}_{l=1}^{L}$, consists of L observations where each observation, $\boldsymbol{b}_l$, corresponds to a random vector of observed bids from an auction. All asymptotic statements are taken to be $L \to \infty$. We assume the researcher observes all bids in an auction, so $\boldsymbol{b}_l = (b_{1l}, ..., b_{p_l l})$ where

[5] We have suppressed the dependence of θ_0 on L for notational convenience.

[6] For example, in the log-normal case $\beta = (\mu, \sigma)$.

p_l denotes the number of participants in the l-th observed auction. Further, we assume the number of participants has support Supp(p) which is a finite subset of $\{2, ..., \overline{p}\}$. Therefore, $\boldsymbol{b}_l$ is a random vector supported on $\cup_{p \in \text{Supp}(p)} \mathbb{R}^p$. Let $\pi(p)$ denote the conditional probability mass function describing the distribution of observed p_l.[7] Lastly, to define the criterion function, let $Q_L(\theta) = m_L(\theta)' m_L(\theta)$ where

$$m_L(\theta) = \frac{1}{L} \sum_{l=1}^{L} \sum_{i=1}^{p_l} \left(\frac{d \log(g(b_{il}, \theta, p_l))}{d\theta} - \mu(\theta, p_l) \right)$$

for each $\theta \in \hat{\Theta}_R$ and $\mu(\theta, p)$ is defined in equation (2.6).

2.4.3 Low-Level Sufficient Conditions

In this section, we provide sufficient conditions for Assumptions (2.1) - (2.5) appearing in section 2.3.3. For the remaining conditions, Assumptions (2.6) and (2.7) depend on the particular choice of test-statistic. To state the assumptions let $\zeta_L \equiv \sup_{v \in [\underline{v}, \overline{v}]} \|P(v)\|$ and $\mathcal{P}$ be the class of distributions satisfying the following assumptions. There exists constants $0 < C, \overline{p} < \infty$ not depending on L or P such that the following assumptions hold.

Assumption 2.8. ***Model Specification*** *The data $\{\boldsymbol{b}_l\}_{l=1}^{L}$ satisfies the following properties*

i) Independent Private Values There exists an independent and identically distributed sequence $\{v_j\}_{j=1}^{n}$ ($n \equiv \sum_{l=1}^{L} p_l$) with v_j supported on $[\underline{v}, \overline{v}]$ with density $f_0(v)$ such that for each $l \in \{1, ..., L\}$, $\boldsymbol{b}_l = (b_{1l}, ..., b_{p_l l})$ with

$$b_{il} = v_{\iota(i,l)} - \frac{\int_{\underline{v}}^{v_{\iota(i,l)}} F_0(t)^{p_l - 1} dt}{F_0(v_{\iota(i,l)})^{p_l - 1}}$$

where $\iota(1,1) = 1, \ \ldots, \ \iota(p_l, 1) = p_l, \ \iota(1,2) = p_l + 1, \ \ldots, \ \iota(p_L, L) = n$.

ii) Distribution of Number of Participants: $\{p_l\}_{l=1}^{L}$ is a sequence of independent and identically distributed random variables with support contained in $\{2, 3, .., \overline{p}\}$ for some $\overline{p} < \infty$.

[7] That is, if $g(\boldsymbol{b}, p)$ denotes the conditional distribution of $\boldsymbol{b}$ (conditional on $\boldsymbol{b} \in \mathbb{R}^p$), $\pi(p) \equiv \int_{\mathbb{R}^p} g(\boldsymbol{b}, p) d\boldsymbol{b}$.

iii) Exogenous Entry: For all $l \in \{1, ...L\}$*,* $v_{il} \perp p_l$ *for all* $i \in \{1, .., p_l\}$*.*

Assumption 2.9. ***Sieve Approximation Error*** *There exists a sequence* $\theta_{0,L} \in \Theta_{K_L}$ *such that*

$$\sup_{v \in [\underline{v}, \overline{v}]} |f(v, \theta_0) - f_0(v)| \le C K_L^{-\gamma} \quad \text{and} \quad \overline{b}(\theta_0, p) \ge \overline{b}_0(p)$$

for some $\gamma > 0$ *where* $\overline{b}_0(p) = \overline{v} - \int_{\underline{v}}^{\overline{v}} F_0(t)^{p-1} dt$ *denotes the largest bid rationalized by* $f_0(v)$*.*

Assumption 2.10. ***Density Bounds*** *For all* $\theta \in \Theta$ *and* $v \in [\underline{v}, \overline{v}]$

$$\frac{1}{C} < f_0(v) \, , \, f(v, \theta) \, , \, f'(v, \theta) \, , \, \phi(v) < C.$$

Assumption 2.11. ***Rate Conditions*** *The following rate conditions hold*

$$\frac{K_L^3 \zeta_L^4}{L} = o_p(\varepsilon_L^6) \, , \, \frac{K_L^4 \zeta_L^2}{L} = o_p(\varepsilon_L^6) \, , \, L K_L^{1-2\gamma} = o_p(\varepsilon_L^2).$$

where ε_L *is the sequence described in section 2.3.3.*

Assumption 2.12. ***Bounded Eigenvalues*** *The eigenvalues of the matrices*

$$\mathbb{E}_0 \left(m(\boldsymbol{b}, \theta_0)' m(\boldsymbol{b}, \theta_0) \right) \quad \text{and} \quad \mathbb{E}_0 \left(\frac{dm(\boldsymbol{b}, \theta_0)}{d\theta'} \right) \mathbb{E}_0 \left(\frac{dm(\boldsymbol{b}, \theta_0)}{d\theta'} \right)'$$

lie in the interval $\left[\frac{1}{C} \, , \, C \right]$

Theorem 2.2. *Assumptions (2.8) - (2.12) imply Assumptions (2.2) - (2.5) in section 2.3.3.*

Remark 1: The proof of the theorem is included in the online supplemental appendix. Assumption (2.8) states the observed bids are generated from a simple first-price auction with independent and private valuations. Assumption (2.9) specifies the rate at which the bias of the sieve converges to zero where γ is controlled by the number of continuous derivatives of $f_0(v)$. Furthermore, this assumption states that the approximating sequence of θ_0 approaches $f_0(v)$ from "inside" the restricted set. Assumption (2.10) 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) Assumption (2.11) is a set of abstract rate conditions governing how quickly/slowly K_L diverges to infinity. 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. Finally, Assumption (2.12) bounds the indicated eigenvalues away from infinity and away from zero. This assumption is the direct analogue of the assumption in standard maximum likelihood estimation assuming information matrix has eigenvalues which are bounded and bounded away from zero.[8]

Remark 2: If $\zeta_L \leq CK_L$, 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 $\frac{K_L^7}{L}$ and $LK_L^{-2(\gamma-1)}$ both converge to zero. The analogous rate conditions for nonparametric regression appearing in Newey (1997) are that $\frac{K_L^6}{L}$ and $LK_L^{-2\gamma}$ converge to zero, so our conditions are slightly stronger. Interestingly, as illustrated in GPV (2000), in order to achieve the optimal rate of convergence for the simple-auction model as achieved 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 γ 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 of $f_0(v)$ to get the same rate conditions as in Newey (1997) for nonparametric regression.

[8] If we were able to interchange the order of integration, $\mathbb{E}_0(m(\boldsymbol{b},\theta_0)'m(\boldsymbol{b},\theta_0))$ is the expectation of the outer-product of the gradient which is equal to $\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)$ which is the negative of the Fisher information matrix. Therefore, in this case, the assumption states the Fisher information matrix has eigenvalues which are bounded and bounded away from zero, which is a nonparametric generalization of the parametric identification condition for maximum likelihood estimators.

2.5 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 (8) 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

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

As practitioners, we only observe the sample $\{\{b_i\}_{i=1}^{p_l}\}_{l=1}^{L}$, 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 $\psi_{\text{rev}}(\theta)$ and 3) the optimal reserve price $r_{\text{opt}}(\theta)$. 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 2.4.1 with $\phi(v)$ as the log-normal density with parameters $\beta = (\beta_\mu, \beta_\sigma)$. We then obtain $P(v)$ by applying the GramSchmidt orthonormalization process to the set of cubic splines with K_L-partitions (i.e. $K_L + 1$ knots and $K_L + 2$ total parameters).[9] 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

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

$$\{1\,,\,x\,,\,\ldots\,,\,x^m\,,\,\max\{x - t_1, 0\}^m\,,\,\ldots\,,\,\max\{x - t_{K_L-1}, 0\}^m\}$$

sample $\{\{b_{il}\}_{i=1}^{p_l}\}_{l=1}^{L}$ we obtain the (parametric) maximum likelihood estimate for $\beta = (\mu, \sigma)$, say $\hat{\beta} = (\hat{\mu}, \hat{\sigma})$. Then, for $j \in \{0, ..., K_L\}$, we obtain the j^{th} knot point as $\Phi^{-1}\big(\frac{j}{K_L}, \hat{\beta}\big)$, 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 sub-model $\phi(v, \beta)$. 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 $\hat{\theta}_L$ according to equation (2.7). When obtaining $\hat{\theta}_{\text{ml}}$ and $\hat{\theta}_{\text{mm}}$, we enforce the constraint $\bar{b}(\theta, p) \geq \max\{b_{il}\}$ in the optimization.[10]

Table (7) reports the root-mean-squared error (RMSE) of the three indicated functionals using three different plug-in estimates : 1) our estimator $\hat{\theta}_L$, 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 $K_L = 3$ knots and the results for the other values of K_L 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 $\hat{\theta}_L$, 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 $\hat{\theta}_L$ as opposed to the non-normal MM and ML

[10]Strictly 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 θ for which the criterion function is not defined.

estimators is small.

Table (8) reports the empirical coverage probabilities of a standard two-sided t-test of the indicated functional at the 5% level. Our proposed confidence bands perform well and have empirical coverage close to the nominal size. The low coverage for $K_L = 1$ 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 $K_L = 4$ for small sample sizes indicates the importance of using a slowly growing sieve in controlling the size of the test.

Table (9) 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_L values we used.[11]

Table (10) 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

[11] All results in Table (9) 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.

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.[12] 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.

2.6 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[13]) and an appraisal on the expected market value of the timber based on the anticipated quality

[12] In fact, over all of our Monte Carlo simulations, none of the confidence sets were intervals using this procedure.

[13] A board-foot is the quantity of wood contained in a one-foot square piece of wood which is one-inch thick.

of the wood.[14] Additionally, the cruise also contains an estimate of the cost of harvesting and manufacturing the timber into a final product.[15]

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.[16] 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), 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.[17]. 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.

[14]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.

[15]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.

[16]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.

[17]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.

The contract is then awarded to the agent that submitted the largest bid in the auction.[18] 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 timber provided by the Forest Service.[19] 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.[20] Specifically,

[18]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.

[19]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.

[20]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

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

$$100,000 \text{ BF}_{\text{Oak}} \frac{\$0.3}{\text{BF}_{\text{Oak}}} \frac{\text{BF}_{\text{med. tree}}}{\$0.4} = 75,000 \text{ BF}_{\text{med. tree}}$$

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.[21] Let π_p an agent's belief that there will be p participants (including the agent) who submit bids in the auction and assume the number of potential participants is bounded.

---

and individual-specific component. For example, such a specification could be

$$v_{il} = X_l \beta + \eta_i$$

where X_l reflects the auction-specific covariates and η_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 β in the construction of confidence sets for functionals of the distribution of η_i. This is an avenue for future research.

[21]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.

It can be shown that the equilibrium bidding function is given by

$$b(v,\theta,\pi) = \sum_p w_p(v,\theta,\pi) b(v,\theta,p) \tag{2.16}$$

where $w_p(v,\theta,\pi) = \pi_p F(v,\theta)^{p-1} / \sum_{\tilde{p}} \pi_{\tilde{p}} F(v,\theta)^{\tilde{p}-1}$ and $b(v,\theta,p)$ is the so-called contingent bidding function given in equation (2.1). For what follows we assume a known value of π.[22] We can use arguments in equation (2.2) 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.[23] Furthermore, as $w_p(\overline{v},\theta,\pi) = \pi_p$, the support of bids depends on the true parameters as

$$\overline{b}(\theta,\pi) = \sum_p \pi_p \overline{b}(\theta,p).$$

Similar to the arguments in section 2.2.2 we can apply Leibniz integral rule to get an identical expression as equation (2.6) using $\overline{b}(\theta,\pi)$ in place of $\overline{b}(\theta,p)$. Similarly, we can get the new value of $\mu(\theta,\pi)$ in the first equality in equation (2.6). With the value $\mu(\theta,\pi)$, all arguments of section 2.3 and the general theory in section 2.3.3 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 π. We take the value of π as fixed, so our confidence bands only incorporate uncertainty in the estimation of θ.[24]

Our final sample of bids consists of 4,458 bids from 1,129 auctions. Figure (9) contains the empirical density of the submitted bids and Table (11) contains distribution of auctions

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

[23] This is given as the reciprocal of the derivative of equation (2.16) with respect to v.

[24] We are currently working on extending the code to jointly estimate (π,θ).

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 timer were measured in thousands of board-feet with no sale restrictions.[25] 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 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.[26]

To estimate the density of valuations, we used quadratic splines with five knot points. The lower-bound of valuations, $\underline{v}$, was set as the smallest observed bid while the upper-bound of the valuations was chosen using a nested-maximum likelihood procedure.[27] 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 β_n.[28,29] Figure (11) displays the estimated density of valuations along with the upper and lower bands of the

[25]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.

[26]The appendix contains more detailed information on the data cleaning.

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

[28]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.

[29]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 α_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 $\frac{\beta_n}{\alpha_n} \to 0$. We use this estimator with $\beta_n = \frac{1}{n^{1/3}}$.

(point-wise) 95% confidence band. Figure 11 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.[30] 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. 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 12 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

[30]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.

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 illusrate this, let $\mathcal{W}(r, \theta)$ 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 $\mathcal{W}(r, \hat{\theta}_n)$. 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

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

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

2.7 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.

Appendix A

Inference under Shape Restrictions

A.1 Tables and Figures

Figure 1: Scatter plots of samples illustrating relation between critical values

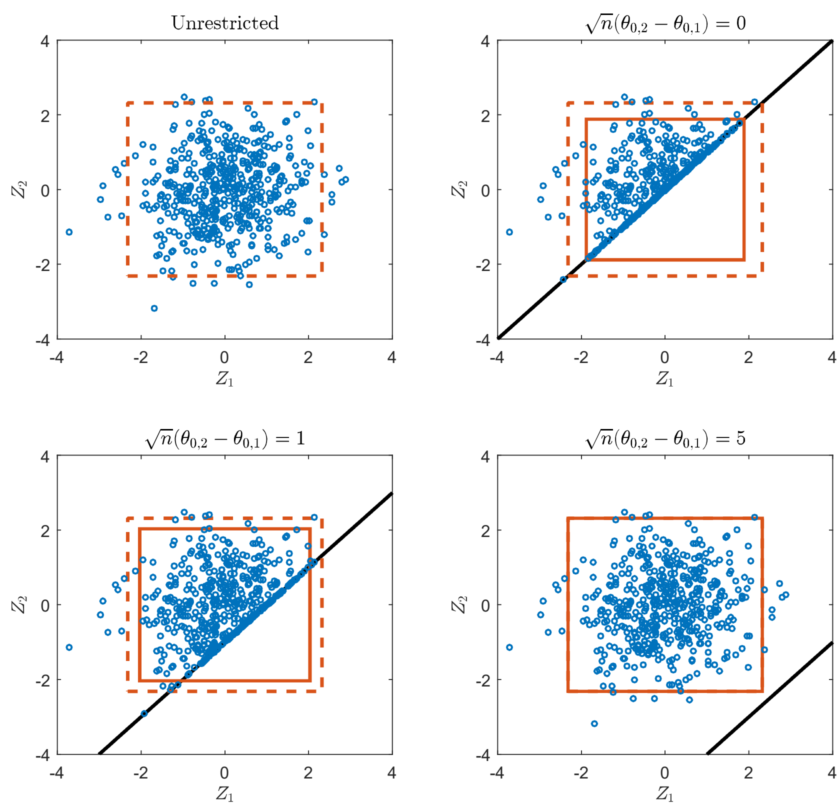


Figure 2: Confidence regions

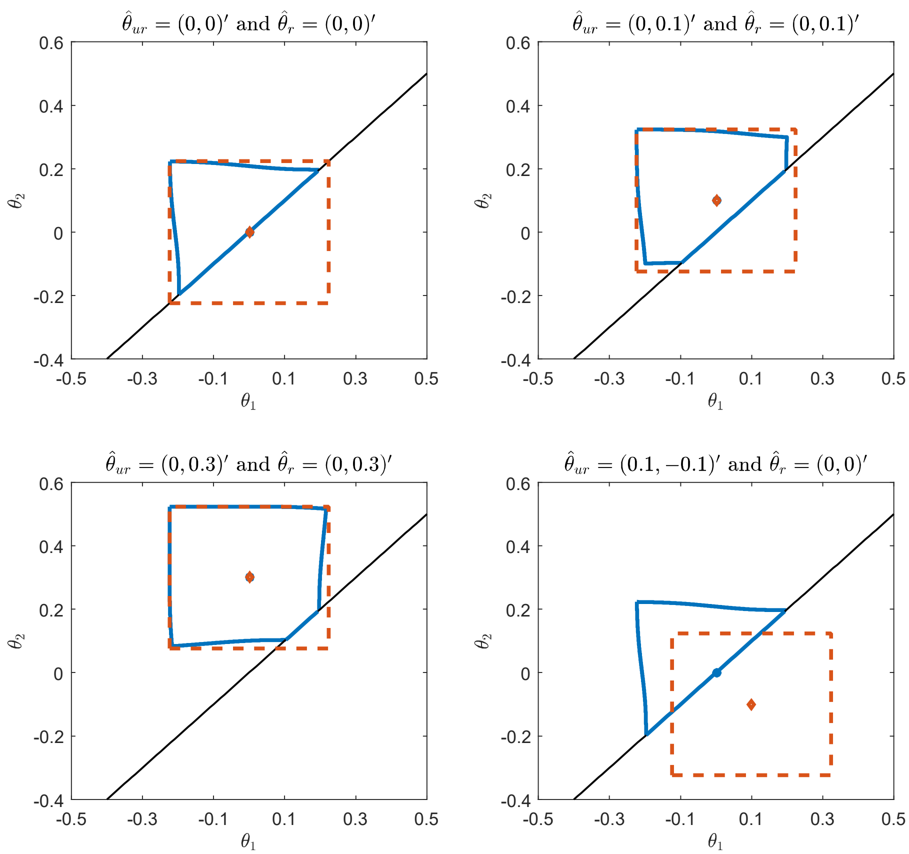


Figure 3: g_0 for different values of c

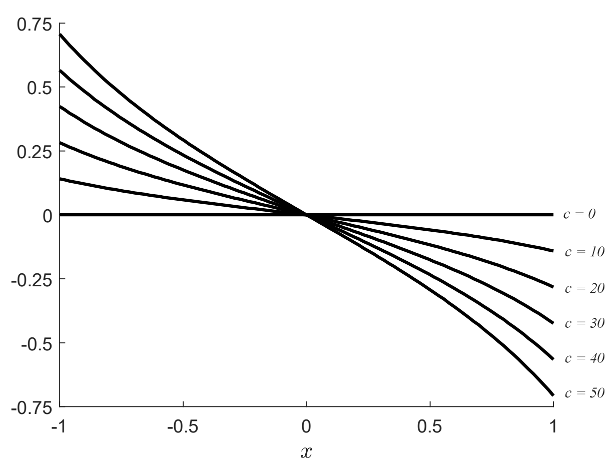


Figure 4: Average confidence bands for NPIV with polynomials and $K_n = 5$

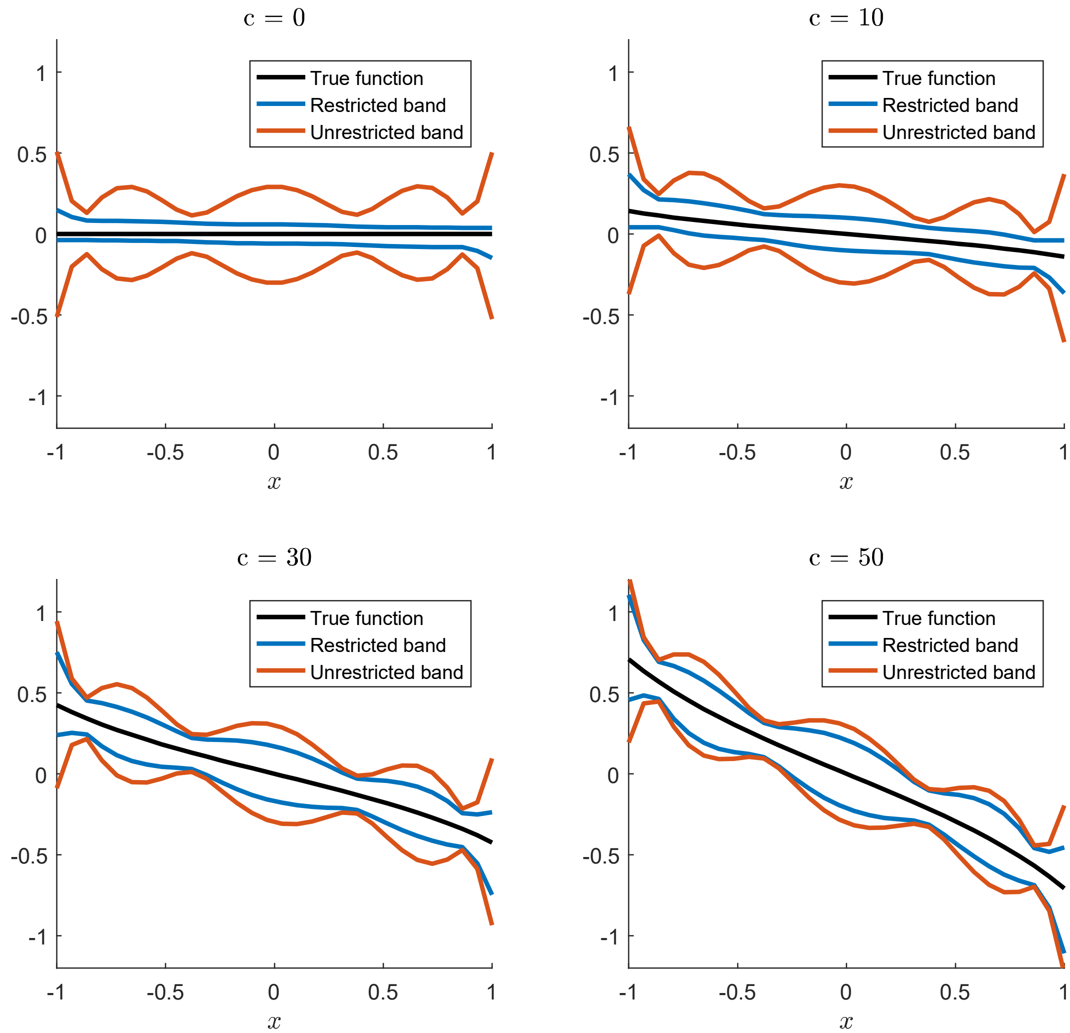


Figure 5: Example confidence bands for NPIV with polynomials and $K_n = 5$

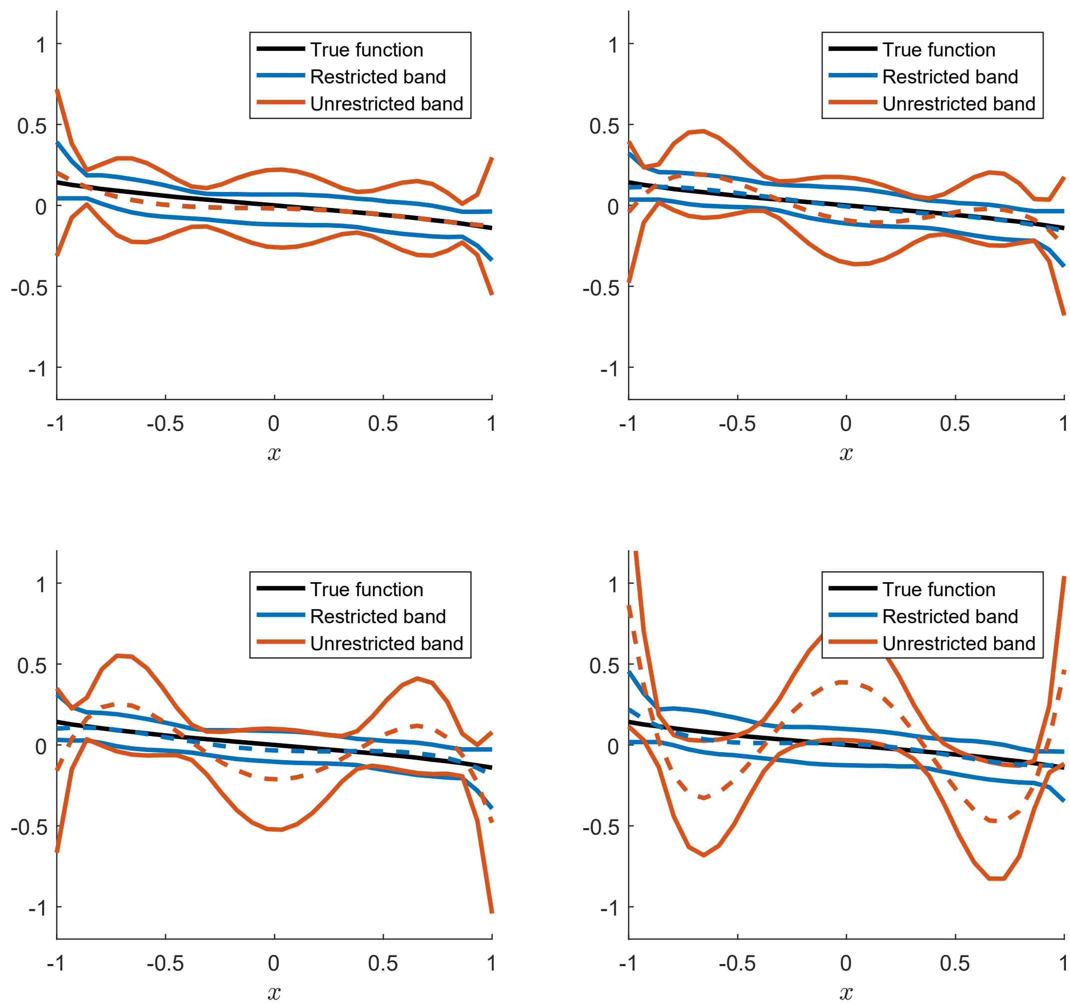
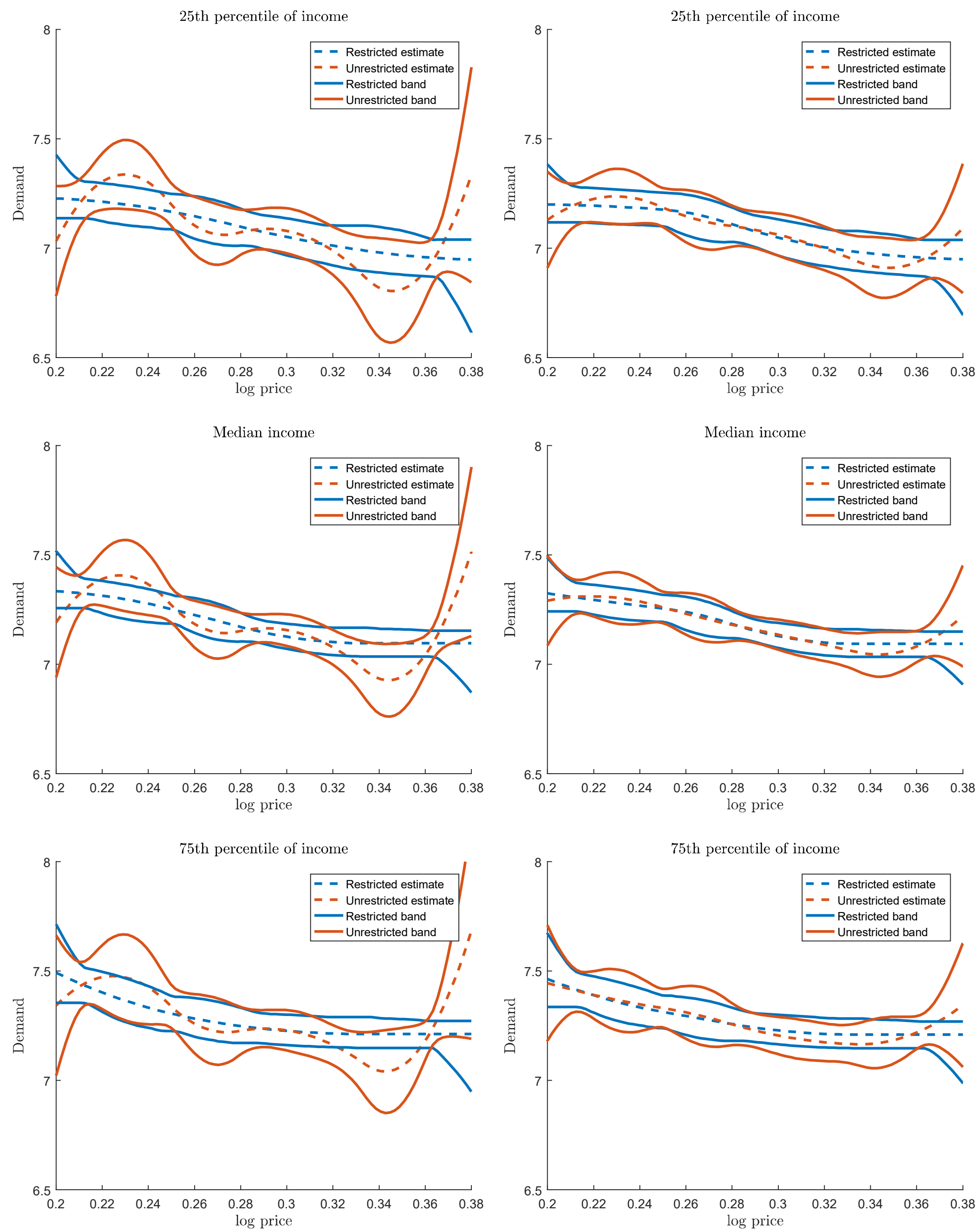


Figure 6: Estimated demand functions



The three figures on the left side use quadratic splines with six knots to construct P_Z. The three figures on the right side use cubic splines with eight knots.

Table 1: Coverage and width comparison for regression with polynomials

K_n	c	cov_{ur}	cov_r	$width_{ur}$	$width_r$	% gains
2	0	0.957	0.948	0.107	0.090	0.175
	2	0.946	0.949	0.107	0.104	0.030
	4	0.939	0.939	0.107	0.106	0.003
	6	0.891	0.891	0.107	0.107	0.000
	8	0.858	0.858	0.107	0.107	0.000
	10	0.813	0.813	0.107	0.107	0.000
3	0	0.949	0.954	0.142	0.109	0.236
	2	0.947	0.963	0.142	0.121	0.143
	4	0.948	0.960	0.142	0.129	0.091
	6	0.925	0.939	0.142	0.134	0.050
	8	0.910	0.910	0.142	0.137	0.028
	10	0.887	0.884	0.142	0.139	0.015
4	0	0.949	0.969	0.172	0.131	0.238
	2	0.946	0.970	0.172	0.146	0.152
	4	0.945	0.969	0.172	0.155	0.097
	6	0.952	0.963	0.172	0.161	0.058
	8	0.930	0.948	0.172	0.166	0.032
	10	0.939	0.947	0.172	0.168	0.018
5	0	0.941	0.970	0.200	0.147	0.262
	2	0.943	0.971	0.200	0.162	0.187
	4	0.945	0.964	0.200	0.173	0.135
	6	0.948	0.960	0.199	0.180	0.097
	8	0.937	0.951	0.200	0.185	0.072
	10	0.948	0.960	0.200	0.189	0.051

Table 2: Coverage and width comparison for NPIV with polynomials

K_n	c	cov_{ur}	cov_r	$width_{ur}$	$width_r$	% gains
2	0	0.946	0.955	0.059	0.046	0.234
	5	0.929	0.929	0.059	0.058	0.016
	10	0.879	0.879	0.060	0.060	0.000
	20	0.608	0.608	0.059	0.059	0.000
	30	0.229	0.229	0.059	0.059	0.000
	40	0.003	0.003	0.059	0.059	0.000
	50	0.000	0.000	0.059	0.059	0.000
3	0	0.933	0.963	0.107	0.061	0.426
	5	0.931	0.949	0.107	0.079	0.257
	10	0.921	0.940	0.107	0.091	0.150
	20	0.821	0.815	0.107	0.101	0.049
	30	0.681	0.680	0.107	0.105	0.018
	40	0.426	0.426	0.107	0.106	0.002
	50	0.201	0.201	0.106	0.106	0.000
4	0	0.951	0.986	0.207	0.092	0.556
	5	0.946	0.982	0.207	0.120	0.422
	10	0.944	0.967	0.208	0.143	0.310
	20	0.942	0.947	0.208	0.171	0.176
	30	0.954	0.967	0.208	0.185	0.103
	40	0.953	0.959	0.207	0.194	0.057
	50	0.952	0.956	0.208	0.199	0.037
5	0	0.959	0.989	0.456	0.122	0.731
	5	0.962	0.994	0.457	0.161	0.649
	10	0.956	0.994	0.460	0.197	0.574
	20	0.957	0.978	0.465	0.248	0.471
	30	0.973	0.985	0.457	0.288	0.377
	40	0.966	0.978	0.462	0.322	0.310
	50	0.953	0.973	0.459	0.345	0.254

Table 3: Width comparison with monotonized bands for NPIV with polynomials

K_n	c	cov_{ur}	cov_r	$widths_{ur}$	$widths_{mon}$	$widths_r$	% empty monotone
5	0	0.959	0.989	0.456	0.210	0.122	0.016
	10	0.956	0.994	0.460	0.299	0.197	0.002
	30	0.973	0.985	0.457	0.373	0.288	0.000
	50	0.953	0.973	0.459	0.411	0.345	0.000

Table 4: Width comparison average derivative

K_n	c	cov_{ur}	cov_r	cov_{cns}	$widths_{ur}$	$widths_{neg}$	$widths_r$	$width_{cns}$	empty neg.	empty CNS
5	-5	0.957	0.000	0.000	0.636	0.208	0.227	0.030	0.118	0.810
	0	0.962	0.957	0.939	0.637	0.325	0.233	0.164	0.026	0.040
	10	0.953	0.967	0.994	0.641	0.593	0.421	0.428	0.000	0.002
	30	0.963	0.971	0.997	0.638	0.638	0.574	0.744	0.000	0.000
	50	0.951	0.963	0.998	0.642	0.642	0.612	0.910	0.000	0.000

A.2 Proofs of Main Results

A.2.1 Non-conservative projections

We now formalize the arguments from Section 1.3.2. Let $h_l(\theta) = c_l + q_l'\theta$, where c_l is a constant and $q_l \in \mathbb{R}^{L_n}$. Let

$$\mathcal{Z}(\hat{\Sigma}) = \left\{ z \in \mathbb{R}^{K_n} : \sup_{l=1,\dots,L_n} \left\{ \left| q_l' z \right| / \sqrt{q_l' \hat{\Sigma} q_l} \right\} \right\},$$

where $c(\hat{\Sigma})$ is such that for $Z \sim N(0, I_{K_n \times K_n})$, $P(\hat{\Sigma}^{1/2} Z \in \mathcal{Z}(\hat{\Sigma}) \mid \hat{\Sigma}) = 1 - \alpha$. We obtain the following corollary.

Corollary A.1. *Suppose that Assumptions 1.1 – 1.6 hold. Let*

$$T(\kappa_n(\hat{\theta}_r - \theta), \hat{\Sigma}) = \sup_{l=1,\dots,L_n} \left\{ \kappa_n \left| q_l' \left(\hat{\theta}_r - \theta \right) \right| / \sqrt{q_l' \hat{\Sigma} q_l} \right\}$$

and let CI be the corresponding confidence region. Let $\Theta_R = \{\theta \in \mathbb{R}^{K_n} : A_n \theta \le b_n\}$. Suppose that, with probability approaching 1, $A_n z < \kappa_n(b_n - A_n\theta)$ for all $\theta \in CI$ and for all $z \in \mathcal{Z}(\hat{\Sigma})$. Then for all $\theta \in CI$, $c_{1-\alpha,n}(\theta, \hat{\Sigma}, \hat{\Omega}) = c(\hat{\Sigma})$ with probability approaching 1 and

$$\lim_{n\to\infty} P\left(\hat{h}_l^L \le h_l(\theta_0) \le \hat{h}_l^U \text{ for all } l = 1, \dots, L_n \right) = 1 - \alpha.$$

Notice that $\kappa_n(b_n - A_n\theta) = \kappa_n(b_n - A_n\theta_0) + \kappa_n A_n(\theta_0 - \theta)$. If θ_0 is sufficiently in the interior of the parameter space, then each element of $\kappa_n(b_n - A_n\theta_0)$ goes to infinity. Moreover, if each element of CI converges in probability to θ_0 at rate κ_n, then each element of $\kappa_n A_n(\theta_0 - \theta)$ is bounded in probability. The condition of the corollary then holds for example if $\mathcal{Z}(\hat{\Sigma})$ is bounded with probability approaching 1, but the condition also allows the set to grow.

Proof of Corollary A.1. Let $z \in \mathbb{R}^{K_n}$ and $\theta \in CI$. Let

$$z_n(\theta, \hat{\Omega}) = \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{K_n} : A_n\lambda \le \kappa_n(b_n - A_n\theta)} \|\lambda - z\|_{\hat{\Omega}}^2.$$

Now notice that if $z \in \mathcal{Z}(\hat{\Sigma})$, then with probability approaching 1 we get $z_n(\theta, \hat{\Omega}) = z$. It therefore follows that $c(\theta, \hat{\Sigma}, \hat{\Omega}) \leq c(\hat{\Sigma})$. Now take $z_n(\theta, \hat{\Omega}) \in \mathcal{Z}(\hat{\Sigma})$. Then by assumption $A_n z_n(\theta, \hat{\Omega}) < \kappa_n(b_n - A_n\theta)$ with probability approaching 1. Since $\hat{\Omega}$ is positive definite with probability approaching 1, it follows that $z_n(\theta, \hat{\Omega}) = z$, because otherwise the projection would be on the boundary of the support. Hence $c(\theta, \hat{\Sigma}, \hat{\Omega}) \geq c(\hat{\Sigma})$ and thus $c(\theta, \hat{\Sigma}, \hat{\Omega}) = c(\hat{\Sigma})$. As shown in Section 1.3.2 if $c(\theta, \hat{\Sigma}, \hat{\Omega}) = c(\hat{\Sigma})$ for all $\theta \in CI$, then the projection is not conservative. □

A.2.2 Useful lemmas

Lemma A.1. *Let Q and $\hat{Q}$ be symmetric and positive definite matrices. Then*

$$\left| \min_{\|v\|=1} v'\hat{Q}v - \min_{\|v\|=1} v'Qv \right| \leq \max_{\|v\|=1} |v'(\hat{Q} - Q)v| \leq \|\hat{Q} - Q\|_S \leq \|\hat{Q} - Q\|$$

and

$$\left| \max_{\|v\|=1} v'\hat{Q}v - \max_{\|v\|=1} v'Qv \right| \leq \max_{\|v\|=1} |v'(\hat{Q} - Q)v| \leq \|\hat{Q} - Q\|_S \leq \|\hat{Q} - Q\|.$$

Proof. For both lines, the first inequality follows from basic properties of minima and maxima. The second and third inequalities follow from the Cauchy-Schwarz inequality. □

Lemma A.2. *Let Q and $\hat{Q}$ be symmetric and positive definite matrices. Then*

$$\|Q^{1/2} - \hat{Q}^{1/2}\|_S \leq \frac{1}{\left(\lambda_{\min}(Q^{1/2}) + \lambda_{\min}(\hat{Q}^{1/2})\right)} \|Q - \hat{Q}\|_S$$

and

$$\|Q - \hat{Q}\|_S \leq \left(\lambda_{\max}(Q^{1/2}) + \lambda_{\max}(\hat{Q}^{1/2})\right) \|Q^{1/2} - \hat{Q}^{1/2}\|_S.$$

Proof. Let λ^2 be the largest eigenvalue of $(Q^{1/2} - \hat{Q}^{1/2})(Q^{1/2} - \hat{Q}^{1/2})$ with unit length eigenvector v_λ. Since $(Q^{1/2} - \hat{Q}^{1/2})$ is symmetric either λ or $-\lambda$ is an eigenvalue of $(Q^{1/2} - \hat{Q}^{1/2})$

with eigenvector v_λ. It follows that

$$\begin{aligned}\sup_{\|v\|=1} |v'(Q-\hat{Q})v| &\geq |v'_\lambda(Q-\hat{Q})v_\lambda| \\ &= |v'_\lambda Q^{1/2}(Q^{1/2}-\hat{Q}^{1/2})v_\lambda + v'_\lambda(Q^{1/2}-\hat{Q}^{1/2})\hat{Q}^{1/2}v_\lambda| \\ &= |\lambda||v'_\lambda Q^{1/2}v_\lambda + v'_\lambda\hat{Q}^{1/2}v_\lambda| \\ &\geq |\lambda|\left(\lambda_{\min}(Q^{1/2}) + \lambda_{\min}(\hat{Q}^{1/2})\right)\end{aligned}$$

and therefore

$$\|Q^{1/2}-\hat{Q}^{1/2}\|_S \leq \frac{1}{\left(\lambda_{\min}(Q^{1/2}) + \lambda_{\min}(\hat{Q}^{1/2})\right)}\|Q-\hat{Q}\|_S.$$

Similarly, for all v with $\|v\|=1$ we have

$$\begin{aligned}\|(Q-\hat{Q})v\| &= \|Q^{1/2}(Q^{1/2}-\hat{Q}^{1/2})v + (Q^{1/2}-\hat{Q}^{1/2})\hat{Q}^{1/2}v\| \\ &\leq \left(\lambda_{\max}(Q^{1/2}) + \lambda_{\max}(\hat{Q}^{1/2})\right)\|Q^{1/2}-\hat{Q}^{1/2}\|_S.\end{aligned}$$

Therefore,

$$\|Q-\hat{Q}\|_S \leq \left(\lambda_{\max}(Q^{1/2}) + \lambda_{\max}(\hat{Q}^{1/2})\right)\|Q^{1/2}-\hat{Q}^{1/2}\|_S.$$

□

A.2.3 Proof of Theorem 1.1

Proof of Theorem 1.1. First notice that $\lambda_{\min}(\Sigma)$ is bounded and bounded away from 0 and since $\|\Sigma-\hat{\Sigma}\|_S \xrightarrow{p} 0$ by Assumption 1.5 it follows from Lemma A.1 that $\lambda_{\min}(\hat{\Sigma})$ is bounded and bounded way from 0 with probability approaching 1. Similarly, $\lambda_{\max}(\Omega)$ is bounded and bounded away from 0 and $\lambda_{\max}(\hat{\Omega})$ is bounded and bounded way from 0 with probability approaching 1. Hence, there exist constants $B_l > 0$ and $B_u < \infty$ such that $B_l \leq \lambda_{\min}(\Sigma), \lambda_{\max}(\Omega) \leq B_u$ and $B_l \leq \lambda_{\min}(\hat{\Sigma}), \lambda_{\max}(\hat{\Omega}) \leq B_u$ with probability approaching 1 uniformly over $P \in \mathcal{P}$.

Also notice that by Assumption 1.3

$$\frac{|\lambda_{\min}(\hat{\Omega}) - \lambda_{\min}(\Omega)|}{\lambda_{\min}(\Omega)} \leq \frac{\|\hat{\Omega} - \Omega\|_S}{\lambda_{\min}(\Omega)} \xrightarrow{p} 0$$

and therefore uniformly over $P \in \mathcal{P}$

$$\left| \frac{\lambda_{\min}(\hat{\Omega})}{\lambda_{\min}(\Omega)} - 1 \right| \xrightarrow{p} 0.$$

Hence $\lambda_{\min}(\hat{\Omega}) > 0$ with probability approaching 1 and, uniformly over $P \in \mathcal{P}$,

$$\left| \frac{\lambda_{\min}(\Omega)}{\lambda_{\min}(\hat{\Omega})} - 1 \right| \xrightarrow{p} 0.$$

Take Z_n as defined in Assumption 1.2 and $\Lambda_n(\theta_0) = \{\lambda \in \mathbb{R}^{K_n} : \lambda = \kappa_n(\theta - \theta_0) \text{ for some } \theta \in \Theta_R\}$ and define

$$Z_n(\theta_0, \Sigma, \Omega) = \underset{\lambda \in \Lambda_n(\theta_0)}{\arg\min} \|\lambda - Z_n\|_{\Omega}^2.$$

By Assumptions 1.5 there exists a constant C such that with probability approaching 1

$$\begin{aligned}
&\left| T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) - T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \right| \\
&\leq \left| T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) - T(Z_n(\theta_0, \Sigma, \Omega), \hat{\Sigma}) \right| + \left| T(Z_n(\theta_0, \Sigma, \Omega), \hat{\Sigma}) - T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \right| \\
&\leq C \left\| \kappa_n(\hat{\theta}_r - \theta_0) - Z_n(\theta_0, \Sigma, \Omega) \right\| + C \|Z_n(\theta_0, \Sigma, \Omega)\| \|\hat{\Sigma} - \Sigma\|_S \\
&\leq C \left\| \kappa_n(\hat{\theta}_r - \theta_0) - Z_n(\theta_0, \Sigma, \hat{\Omega}) \right\| + C \left\| Z_n(\theta_0, \Sigma, \Omega) - Z_n(\theta_0, \Sigma, \hat{\Omega}) \right\| \\
&\qquad + C \|Z_n(\theta_0, \Sigma, \Omega)\| \|\hat{\Sigma} - \Sigma\|_S.
\end{aligned}$$

We now first prove that each term on the right hand side is $o_p(\varepsilon_n)$ uniformly over $P \in \mathcal{P}$.

Since $\Lambda_n(\theta_0)$ is closed and convex it follows from Assumptions 1.1 and 1.2 that

$$\begin{aligned}
&\left\|\kappa_n(\hat{\theta}_r - \theta_0) - Z_n(\theta_0, \Sigma, \hat{\Omega})\right\| \\
&\qquad \leq \left\|\underset{\lambda\in\Lambda_n(\theta_0)}{\arg\min} \|\lambda - \kappa_n(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}} - Z_n(\theta_0, \Sigma, \hat{\Omega})\right\| + \|R_n\| \\
&\qquad \leq \lambda_{\min}(\hat{\Omega})^{-1/2} \left\|\underset{\lambda\in\Lambda_n(\theta_0)}{\arg\min} \|\lambda - \kappa_n(\hat{\theta}_{ur} - \theta_0)\|^2_{\hat{\Omega}} - Z_n(\theta_0, \Sigma, \hat{\Omega})\right\|_{\hat{\Omega}} + \|R_n\| \\
&\qquad \leq \lambda_{\min}(\hat{\Omega})^{-1/2} \|\kappa_n(\hat{\theta}_{ur} - \theta_0) - Z_n\|_{\hat{\Omega}} + \|R_n\| \\
&\qquad \leq \sqrt{\frac{\lambda_{\max}(\hat{\Omega})}{\lambda_{\min}(\hat{\Omega})}} \|\kappa_n(\hat{\theta}_{ur} - \theta_0) - Z_n\| + \|R_n\|.
\end{aligned}$$

Also notice that $\sqrt{\lambda_{\max}(\hat{\Omega})} = O_p(1)$ and $\left|\frac{\lambda_{\min}(\Omega)}{\lambda_{\min}(\hat{\Omega})} - 1\right| = o_p(1)$ uniformly over $P \in \mathcal{P}$. Combined with Assumptions 1.1 and 1.2 this implies that

$$C\left\|\kappa_n(\hat{\theta}_r - \theta_0) - Z_n(\theta_0, \Sigma, \hat{\Omega})\right\| = o_p(\varepsilon_n)$$

uniformly over $P \in \mathcal{P}$.

Next notice that the $K_n \times 1$ zero vector is in $\Lambda_n(\theta_0)$. Therefore

$$\|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|_\Omega \leq \|Z_n\|_\Omega$$

and thus,

$$\sqrt{\lambda_{\min}(\Omega)}\|Z_n(\theta_0, \Sigma, \Omega) - Z_n\| \leq \sqrt{\lambda_{\max}(\Omega)}\|Z_n\|.$$

It follows that

$$\begin{aligned}
\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n\|^2_{\hat{\Omega}} &\leq \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|^2_{\hat{\Omega}} \\
&= \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|^2_{\Omega} + \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|^2_{\hat{\Omega}-\Omega} \\
&\leq \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|^2_{\Omega} + \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|^2\|\hat{\Omega} - \Omega\|_S \\
&\leq \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|^2_{\Omega} + \frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)}\|Z_n\|^2\|\hat{\Omega} - \Omega\|_S.
\end{aligned}$$

Let

$$\hat{V}_1 = \frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)}\|Z_n\|^2\|\hat{\Omega} - \Omega\|_S.$$

Analogously, we get

$$\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n\|_\Omega^2 \quad \leq \quad \|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n\|_{\hat{\Omega}}^2 + \hat{V}_2,$$

where

$$\hat{V}_2 = \frac{\lambda_{\max}(\hat{\Omega})}{\lambda_{\min}(\hat{\Omega})} \, \|Z_n\|^2 \|\hat{\Omega} - \Omega\|_S.$$

Since $\Lambda_n(\theta_0)$ is convex it follows that for any $\gamma \in (0, 1)$

$$\begin{aligned}
\|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|_\Omega^2 &\leq \|\gamma Z_n(\theta_0, \Sigma, \Omega) + (1-\gamma) Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n\|_\Omega^2 \\
&= \gamma \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|_\Omega^2 + (1-\gamma)\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n\|_\Omega^2 \\
&\quad -\gamma(1-\gamma)\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n(\theta_0, \Sigma, \Omega)\|_\Omega^2 \\
&\leq \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|_\Omega^2 + (1-\gamma)(\hat{V}_1 + \hat{V}_2) \\
&\quad -\lambda_{\min}(\Omega)\gamma(1-\gamma)\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n(\theta_0, \Sigma, \Omega)\|^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n(\theta_0, \Sigma, \Omega)\|^2 \\
&\qquad \leq \frac{1}{\lambda_{\min}(\Omega)\gamma}(\hat{V}_1 + \hat{V}_2) \\
&\qquad = \frac{1}{\lambda_{\min}(\Omega)\gamma}\left(\frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)} + \frac{\lambda_{\max}(\hat{\Omega})}{\lambda_{\min}(\hat{\Omega})}\right)\|Z_n\|^2\|\hat{\Omega} - \Omega\|_S \\
&\qquad \leq \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)\gamma}\left(\frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)} + \frac{\lambda_{\max}(\hat{\Omega})}{\lambda_{\min}(\hat{\Omega})}\right)\|\Sigma^{-1/2} Z_n\|^2\|\hat{\Omega} - \Omega\|_S.
\end{aligned}$$

Since

$$\left|\frac{\lambda_{\min}(\Omega)}{\lambda_{\min}(\hat{\Omega})} - 1\right| \xrightarrow{p} 0,$$

$\lambda_{\max}(\hat{\Omega})$ is bounded with probability approaching 1, and $\|\Sigma^{-1/2} Z_n\|^2 = O_p(K_n)$ by Markov's inequality, it follows from Assumption 1.3 that

$$C^2 \|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n(\theta_0, \Sigma, \Omega)\|^2 = o_p(\varepsilon_n^2)$$

uniformly over $P \in \mathcal{P}$ and thus

$$C\|Z_n(\theta_0, \Sigma, \hat{\Omega}) - Z_n(\theta_0, \Sigma, \Omega)\| = o_p(\varepsilon_n)$$

uniformly over $P \in \mathcal{P}$.

From the arguments above and the reverse triangle inequality we have

$$\|Z_n(\theta_0, \Sigma, \Omega)\|_\Omega - \|Z_n\|_\Omega \leq \|Z_n(\theta_0, \Sigma, \Omega) - Z_n\|_\Omega \leq \|Z_n\|_\Omega$$

and therefore

$$C\|Z_n(\theta_0, \Sigma, \Omega)\|\|\hat{\Sigma} - \Sigma\|_S \leq 2C\|\hat{\Sigma} - \Sigma\|_S \sqrt{\frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)}}\|Z_n\|$$

and by Assumption 1.3 and $\|Z_n\| = O_p(\sqrt{\lambda_{\max}(\Sigma)K_n})$

$$C\|Z_n(\theta_0, \Sigma, \Omega)\|\|\hat{\Sigma} - \Sigma\|_S = o_p(\varepsilon_n)$$

uniformly over $P \in \mathcal{P}$.

Next define

$$\begin{aligned} B_n &= C\left\|\kappa_n(\hat{\theta}_r - \theta_0) - Z_n(\theta_0, \Sigma, \hat{\Omega})\right\| + C\left\|Z_n(\theta_0, \Sigma, \Omega) - Z_n(\theta_0, \Sigma, \hat{\Omega})\right\| \\ &\quad + C\|Z_n(\theta_0, \Sigma, \Omega)\|\|\hat{\Sigma} - \Sigma\|_S, \end{aligned}$$

The previous derivations imply that

$$\sup_{P \in \mathcal{P}} P\left(B_n \geq \frac{1}{2}\varepsilon_n\right) \to 0.$$

Therefore

$$\begin{aligned} & P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})\right) \\ & \quad \geq P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) - B_n\right) - o(1) \\ & \quad \geq P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) - \frac{1}{2}\varepsilon_n, B_n \leq \frac{1}{2}\varepsilon_n\right) - o(1) \\ & \quad \geq P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) - \frac{1}{2}\varepsilon_n\right) - P\left(B_n \geq \frac{1}{2}\varepsilon_n\right) - o(1), \end{aligned}$$

where the $o(1)$ term belongs to $P(B_l \leq \lambda_{\min}(\hat{\Sigma}) \leq B_u, B_l \leq \lambda_{\max}(\hat{\Omega}) \leq B_u)$ and it converges to 0 uniformly over $P \in \mathcal{P}$. Similarly, we get

$$\begin{aligned} &P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})\right) \\ &\quad \leq P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})) + \frac{1}{2}\varepsilon_n\right) + P\left(B_n \geq \frac{1}{2}\varepsilon_n\right) + o(1). \end{aligned}$$

We next show that for any sufficiently small $\delta_q \in (0, \alpha)$ it holds that

$$\sup_{P \in \mathcal{P}} \left| P\left(c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) \geq c_{1-\alpha-\delta_q,n}(\theta_0, \Sigma, \Omega) - \frac{1}{2}\varepsilon_n\right) - 1\right| \to 0$$

and

$$\sup_{P \in \mathcal{P}} \left| P\left(c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) \leq c_{1-\alpha+\delta_q,n}(\theta_0, \Sigma, \Omega) + \frac{1}{2}\varepsilon_n\right) - 1\right| \to 0.$$

It then follows that

$$\begin{aligned} &\inf_{P \in \mathcal{P}} P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \\ &\quad \geq \inf_{P \in \mathcal{P}} P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha-\delta_q,n}(\theta_0, \Sigma, \Omega)\right) - o(1) \end{aligned}$$

which implies that

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \geq 1 - \alpha - \delta_q.$$

Since δ_q was arbitrary

$$\liminf_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega}) + \varepsilon_n\right) \geq 1 - \alpha,$$

which is the first conclusion of Theorem 1.1. Similarly, for all δ_q sufficiently small

$$\begin{aligned} &P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})\right) \\ &\quad \geq P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha-\delta_q,n}(\theta_0, \Sigma, \Omega) - \varepsilon_n\right) - o(1) \end{aligned}$$

which implies that

$$\begin{aligned} &P\left(T(\kappa_n(\hat{\theta}_r - \theta_0), \hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0, \hat{\Sigma}, \hat{\Omega})\right) - (1 - \alpha) \\ &\quad \geq P\left(T(Z_n(\theta_0, \Sigma, \Omega), \Sigma) \leq c_{1-\alpha-\delta_q,n}(\theta_0, \Sigma, \Omega) - \varepsilon_n\right) - (1 - \alpha - \delta_q) - \delta_q - o(1). \end{aligned}$$

Analogously,

$$
\begin{aligned}
&P\left(T(\kappa_n(\hat{\theta}_r-\theta_0),\hat{\Sigma}) \le c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})\right)-(1-\alpha)\\
&\quad\le\ P\left(T(Z_n(\theta_0,\Sigma,\Omega),\Sigma)\le c_{1-\alpha+\delta_q,n}(\theta_0,\Sigma,\Omega)+\varepsilon_n\right)-(1-\alpha+\delta_q)+\delta_q+o(1).
\end{aligned}
$$

Hence, if Assumption 1.6 holds, then for all $\delta_q \in (0,\alpha)$

$$
\limsup_{n\to\infty}\sup_{P\in\mathcal{P}}\left|P\left(T(\kappa_n(\hat{\theta}_r-\theta_0),\hat{\Sigma}) \le c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})\right)-(1-\alpha)\right|\le\delta_q
$$

and since δ_q was arbitrary

$$
\sup_{P\in\mathcal{P}}\left|P\left(T(\kappa_n(\hat{\theta}_r-\theta_0),\hat{\Sigma}) \le c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})\right)-(1-\alpha)\right|\to 0.
$$

For the final step of the proof let $\delta_q \in (0,\alpha)$ be arbitrary. Let $\delta_\varepsilon > 0$, which may depend on δ_q, and define the set $\mathcal{H}_n$ as all $(\tilde{\Sigma},\tilde{\Omega})$ on the support of $(\hat{\Sigma},\hat{\Omega})$ such that $B_l \le \lambda_{\min}(\tilde{\Sigma}),\lambda_{\max}(\tilde{\Omega})\le B_u$,

$$
\sqrt{K_n}\frac{\sqrt{\lambda_{\max}(\Sigma)}}{\sqrt{\lambda_{\min}(\Omega)}}\|\tilde{\Sigma}-\Sigma\|_S\le\delta_\varepsilon\varepsilon_n
$$

and

$$
K_n\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\left(\frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)}+\frac{\lambda_{\max}(\tilde{\Omega})}{\lambda_{\min}(\tilde{\Omega})}\right)\|\tilde{\Omega}-\Omega\|_S\le\delta_\varepsilon^2\varepsilon_n^2.
$$

Notice that Assumption 1.3 implies that

$$
\sup_{P\in\mathcal{P}}\left|P\left((\hat{\Sigma},\hat{\Omega})\in\mathcal{H}_n\right)-1\right|\to 0.
$$

Let $\tilde{Z}_n \sim N(0, I_{K_n\times K_n})$ be independent of $\hat{\Sigma}$ and $\hat{\Omega}$ and define

$$
\tilde{Z}_n(\theta_0,\Sigma,\Omega)=\underset{\lambda\in\Lambda_n(\theta_0)}{\arg\min}\,\|\lambda-\Sigma^{1/2}\tilde{Z}_n\|_\Omega^2.
$$

For any $(\Sigma^*, \Omega^*) \in \mathcal{H}_n$ we get by Assumption 1.5 that

$$\begin{aligned}
&\left|T(\tilde{Z}_n(\theta_0, \Sigma^*, \Omega^*), \Sigma^*) - T(\tilde{Z}_n(\theta_0, \Sigma, \Omega), \Sigma)\right| \\
&\quad \leq \left|T(\tilde{Z}_n(\theta_0, \Sigma^*, \Omega^*), \Sigma^*) - T(\tilde{Z}_n(\theta_0, \Sigma, \Omega), \Sigma^*)\right| \\
&\qquad + \left|T(\tilde{Z}_n(\theta_0, \Sigma, \Omega), \Sigma^*) - T(\tilde{Z}_n(\theta_0, \Sigma, \Omega), \Sigma)\right| \\
&\quad \leq C \left\|\tilde{Z}_n(\theta_0, \Sigma^*, \Omega^*) - \tilde{Z}_n(\theta_0, \Sigma, \Omega^*)\right\| \\
&\qquad + C \left\|\tilde{Z}_n(\theta_0, \Sigma, \Omega) - \tilde{Z}_n(\theta_0, \Sigma, \Omega^*)\right\| + C\|\tilde{Z}_n(\theta_0, \Sigma, \Omega)\|\|\Sigma - \Sigma^*\|_S.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\|\tilde{Z}_n(\theta_0, \Sigma^*, \Omega^*) - \tilde{Z}_n(\theta_0, \Sigma, \Omega^*)\| &\leq \sqrt{\lambda_{\max}(\Omega^*)}\|(\Sigma^*)^{1/2}\tilde{Z}_n - \Sigma^{1/2}\tilde{Z}_n\| \\
&\leq \sqrt{\lambda_{\max}(\Omega^*)}\|(\Sigma^*)^{1/2} - \Sigma^{1/2}\|_S\|\tilde{Z}_n\| \\
&\leq \frac{\sqrt{\lambda_{\max}(\Omega^*)}}{\sqrt{\lambda_{\min}(\Sigma)} + \sqrt{\lambda_{\min}(\Sigma^*)}}\|\Sigma^* - \Sigma\|_S\|\tilde{Z}_n\|,
\end{aligned}$$

where the last line follows from Lemma A.2. Also by the previous results

$$\begin{aligned}
&\|\tilde{Z}_n(\theta_0, \Sigma, \Omega^*) - \tilde{Z}_n(\theta_0, \Sigma, \Omega)\|^2 \\
&\quad \leq \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)\gamma}\left(\frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)} + \frac{\lambda_{\max}(\Omega^*)}{\lambda_{\min}(\Omega^*)}\right)\|\tilde{Z}_n\|^2\|\Omega^* - \Omega\|_S.
\end{aligned}$$

and

$$\|Z_n(\theta_0, \Sigma, \Omega)\| \leq 2\sqrt{\frac{\lambda_{\max}(\Omega)\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}}\|\tilde{Z}_n\|.$$

Let

$$\begin{aligned}
H_n &= C\left\|\tilde{Z}_n(\theta_0, \Sigma^*, \Omega^*) - Z_n(\theta_0, \Sigma, \Omega^*)\right\| + C\left\|Z_n(\theta_0, \Sigma, \Omega) - Z_n(\theta_0, \Sigma, \Omega^*)\right\| \\
&\quad + C\|Z_n(\theta_0, \Sigma, \Omega)\|\|\Sigma - \Sigma^*\|_S.
\end{aligned}$$

Then, for constants M_1 and M_2 that do not depend on P or δ_ε

$$\begin{aligned} H_n^2 &\leq 4C^2 \left\|\tilde{Z}_n(\theta_0,\Sigma^*,\Omega^*) - Z_n(\theta_0,\Sigma,\Omega^*)\right\|^2 + 4C^2\|Z_n(\theta_0,\Sigma,\Omega)\|^2\|\Sigma-\Sigma^*\|_S^2 \\ &\quad +4C^2\left\|Z_n(\theta_0,\Sigma,\Omega) - Z_n(\theta_0,\Sigma,\Omega^*)\right\|^2 \\ &\leq M_1\|\Sigma-\Sigma^*\|_S^2\|\tilde{Z}_n\|^2 + M_1\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\tilde{Z}_n\|^2\|\Sigma-\Sigma^*\|_S^2 \\ &\quad +M_1\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\left(\frac{\lambda_{\max}(\Omega)}{\lambda_{\min}(\Omega)} + \frac{\lambda_{\max}(\Omega^*)}{\lambda_{\min}(\Omega^*)}\right)\|\tilde{Z}_n\|^2\|\Omega^*-\Omega\|_S \\ &\leq M_2\varepsilon_n^2\delta_\varepsilon^2\frac{\|\tilde{Z}_n\|^2}{K_n}. \end{aligned}$$

Since $E(\|\tilde{Z}_n\|^2) = K_n$ it follows from Markov's inequality that

$$\sup_{P\in\mathcal{P}} P\left(H_n \geq \frac{1}{2}\varepsilon_n\right) \leq 4M_2\delta_\varepsilon^2.$$

Therefore

$$\begin{aligned} 1-\alpha &= P\left(T(\tilde{Z}_n(\theta_0,\Sigma^*,\Omega^*)) \leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*)\right) \\ &\leq P\left(T(\tilde{Z}_n(\theta_0,\Sigma,\Omega)) \leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*) + H_n\right) \\ &\leq P\left(T(\tilde{Z}_n(\theta_0,\Sigma,\Omega)) \leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*) + \frac{1}{2}\varepsilon_n\right) + 4M_2\delta_\varepsilon^2. \end{aligned}$$

It follows that we can pick δ_ε such that for any $P \in \mathcal{P}$, and any $(\Sigma^*,\Omega^*) \in \mathcal{H}_n$

$$1-\alpha-\delta_q \leq P\left(T(\tilde{Z}_n(\theta_0,\Sigma,\Omega)) \leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*) + \frac{1}{2}\varepsilon_n\right)$$

and thus

$$c_{1-\alpha-\delta_q,n}(\theta_0,\Sigma,\Omega) \leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*) + \frac{1}{2}\varepsilon_n.$$

Hence

$$P\left((\hat{\Sigma},\hat{\Omega}) \in \mathcal{H}_n\right) \leq P\left(c_{1-\alpha-\delta_q,n}(\theta_0,\Sigma,\Omega) \leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega}) + \frac{1}{2}\varepsilon_n\right)$$

and since

$$\sup_{P\in\mathcal{P}}\left|P\left((\hat{\Sigma},\hat{\Omega}) \in \mathcal{H}_n\right) - 1\right| \to 0$$

we have

$$\sup_{P\in\mathcal{P}}\left|P\left(c_{1-\alpha-\delta_q,n}(\theta_0,\Sigma,\Omega)\leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})+\frac{1}{2}\varepsilon_n\right)-1\right|\rightarrow 0.$$

Analogously, for any $(\Sigma^*,\Omega^*)\in\mathcal{H}_n$

$$\begin{aligned}1-\alpha &= P\left(T(\tilde{Z}_n(\theta_0,\Sigma^*,\Omega^*))\leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*)\right)\\ &\geq P\left(T(\tilde{Z}_n(\theta_0,\Sigma,\Omega))\leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*)-H_n\right)\\ &\geq P\left(T(\tilde{Z}_n(\theta_0,\Sigma,\Omega))\leq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*)-\frac{1}{2}\varepsilon_n\right)-4M_2\delta_\varepsilon^2.\end{aligned}$$

It follows that we can pick δ_ε small enough such that for any $P\in\mathcal{P}$, and any $(\Sigma^*,\Omega^*)\in\mathcal{H}_n$

$$c_{1-\alpha+\delta_q,n}(\theta_0,\Sigma,\Omega)\geq c_{1-\alpha,n}(\theta_0,\Sigma^*,\Omega^*)-\frac{1}{2}\varepsilon_n.$$

Hence

$$P\left((\hat{\Sigma},\hat{\Omega})\in\mathcal{H}_n\right)\leq P\left(c_{1-\alpha+\delta_q,n}(\theta_0,\Sigma,\Omega)\geq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})-\frac{1}{2}\varepsilon_n\right)$$

and

$$\sup_{P\in\mathcal{P}}\left|P\left(c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})\leq c_{1-\alpha+\delta_q,n}(\theta_0,\Sigma,\Omega)+\frac{1}{2}\varepsilon_n\right)-1\right|\rightarrow 0.$$

□

A.2.4 Proofs of results from Sections 1.4 and 1.5

Proof of Theorem 1.2. We prove the theorem by verifying the assumptions of Theorem 1.1 with $\kappa_n=\sqrt{n}$. Before we do so notice that $\hat{p}(x_k)-p(x_k)=o_p(1)$ and $\hat{\sigma}^2(x_k)-\sigma^2(x_k)=o_p(1)$ uniformly over $P\in\mathcal{P}$. Therefore, $\frac{\hat{\sigma}^2(x_k)}{\hat{p}(x_k)}-\frac{\sigma^2(x_k)}{p(x_k)}=o_p(1)$ and since we assume that K is fixed and that $\sigma^2(x)$ and $p(x)$ are bounded away from 0, also

$$\max_{k=1,\dots,K}\left|\frac{\hat{\sigma}^2(x_k)}{\hat{p}(x_k)}-\frac{\sigma^2(x_k)}{p(x_k)}\right|\xrightarrow{p}0\quad\text{and}\quad\max_{k=1,\dots,K}\left|\frac{\hat{p}(x_k)}{\hat{\sigma}^2(x_k)}-\frac{p(x_k)}{\sigma^2(x_k)}\right|\xrightarrow{p}0.$$

(1.1) Assumption 1.1 holds by construction with $\hat{\Omega}=\hat{\Sigma}^{-1}$ and $\|R_n\|=0$.

(1.2) To verify Assumption 1.2, we apply Yurinskii's coupling. See for example Appendix D1 of Chernozhukov et al. (2013). To do so, first define P_X as the $n \times K$ matrix with element (i,k) equal to $\mathbf{1}(X_i = x_k)$ and denote the ith row by P'_{Xi}. We apply the coupling argument to $\frac{1}{\sqrt{n}} P'_X U$. The theorem implies that there is $W \sim N(0, diag(\sigma^2(x_1)p(x_1), \ldots, \sigma^2(x_K)p(x_K)))$ such that for any $\delta > 0$ and $\varepsilon_n > 0$

$$P\left(\left\| \frac{1}{\sqrt{n}} P'_X U - W \right\| \geq 3\delta\varepsilon_n \right) \leq C_0 \left(\frac{K}{n^{1/2}} \frac{M}{\delta^3 \varepsilon_n^3} \right) \left(1 + \frac{\left| \log\left(\frac{K}{n^{1/2}} \frac{M}{\delta^3\varepsilon_n^3} \right) \right|}{K} \right),$$

where

$$\begin{aligned} M &= E[\|(U\mathbf{1}(X = x_1), \ldots, U\mathbf{1}(X = x_K))'\|^3] \\ &= K^{3/2} E\left[\left(\frac{1}{K} \sum_{k=1}^{K} (U\mathbf{1}(X = x_k))^2 \right)^{3/2} \right] \\ &\leq K^{3/2} E\left[\left(\frac{1}{K} \sum_{k=1}^{K} |U\mathbf{1}(X = x_k)|^3 \right) \right] \\ &= K^{1/2} \sum_{k=1}^{K} E\left[|U|^3 \mid X = x_k \right] p(x_k). \end{aligned}$$

Since M is bounded, it follows that for any $\varepsilon_n > 0$ with $\frac{1}{\sqrt{n}} = o(\varepsilon_n^3)$ and any $\delta > 0$

$$\sup_{P \in \mathcal{P}} P\left(\left\| \frac{1}{\sqrt{n}} P'_X U - W \right\| \geq \delta\varepsilon_n \right) \to 0.$$

Let $U = (U_1, \ldots, U_n)$. Now write

$$\begin{aligned} &\|\sqrt{n}(\hat{\theta}_{ur} - \theta_0) - (E(P_{Xi}P'_{Xi}))^{-1} W\| \\ &= \left\| ((1/n)P'_X P_X)^{-1} \frac{1}{\sqrt{n}} P'_X U - (E(P_{Xi}P'_{Xi}))^{-1} W \right\| \\ &\leq \left\| ((1/n)P'_X P_X)^{-1} \left(\frac{1}{\sqrt{n}} P'_X U - W \right) \right\| + \left\| \left(((1/n)P'_X P_X)^{-1} - (E(P_{Xi}P'_{Xi}))^{-1} \right) W \right\| \\ &\leq \max_{k=1,\ldots,K} (\hat{p}(x_k))^{-1} \left\| \left(\frac{1}{\sqrt{n}} P'U - W \right) \right\| + \max_{k=1,\ldots,K} |\hat{p}(x_k)^{-1} - p(x_k)^{-1}| \|W\|. \end{aligned}$$

By the previous results, it follows that for any $\varepsilon_n > 0$ with $\frac{1}{\sqrt{n}} = o(\varepsilon_n^3)$ and any $\delta > 0$

$$\sup_{P \in \mathcal{P}} P(\|\sqrt{n}(\hat{\theta}_{ur} - \theta_0) - (E(P_{Xi}P'_{Xi}))^{-1} W\| \geq \delta\varepsilon_n) \to 0$$

or

$$\sup_{P\in\mathcal{P}} P(\|\sqrt{n}(\hat{\theta}_{ur}-\theta_0)-Z_n\| \geq \delta\varepsilon_n) \to 0$$

where $Z_n \sim N(0,\Sigma)$. The result now follows because $\Omega = \Sigma^{-1}$ and $\lambda_{\max}(\Sigma) \leq C^2$.

(1.3) We have

$$\|\hat{\Omega}-\Omega\|_S = \max_{k=1,\ldots,K}\left|\frac{\hat{p}(x_k)}{\hat{\sigma}^2(x_k)} - \frac{p(x_k)}{\sigma^2(x_k)}\right| = O_p(1/\sqrt{n})$$

and

$$\|\hat{\Sigma}-\Sigma\|_S = \max_{k=1,\ldots,K}\left|\frac{\hat{\sigma}^2(x_k)}{\hat{p}(x_k)} - \frac{\sigma^2(x_k)}{p(x_k)}\right| = O_p(1/\sqrt{n})$$

uniformly over $P \in \mathcal{P}$. Moreover, $\lambda_{\max}(\Sigma)$ and $\lambda_{\max}(\Omega)$ are bounded away from 0. Since $\frac{1}{\sqrt{n}} = o(\varepsilon_n^3)$, Assumption 1.3 holds.

(1.4) Assumption 1.4 holds by assumption.

(1.5) By the reverse triangle inequality

$$\begin{aligned}
|T(z,\Sigma)-T(w,\Sigma)| &\leq T(z-w,\Sigma) \\
&= \max\left\{\left|\frac{z_1-w_1}{\sqrt{\Sigma_{11}}}\right|,\ldots,\left|\frac{z_K-w_K}{\sqrt{\Sigma_{KK}}}\right|\right\} \\
&\leq \sqrt{\sum_{k=1}^{K}\left(\frac{z_k-w_k}{\sqrt{\Sigma_{kk}}}\right)^2} \\
&\leq \max_{k=1,\ldots,K}\left(\Sigma_{kk}\right)^{-1/2}\|z-w\| \\
&= \lambda_{\max}(\Sigma^{-1/2})\|z-w\| \\
&= \frac{1}{\sqrt{\lambda_{\min}(\Sigma)}}\|z-w\|.
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&|T(z,\Sigma) - T(z,\tilde{\Sigma})| \\
&\qquad \leq \max\left\{\left|\frac{z_1}{\sqrt{\Sigma_{11}}} - \frac{z_1}{\sqrt{\tilde{\Sigma}_{11}}}\right|, \ldots, \left|\frac{z_K}{\sqrt{\Sigma_{KK}}} - \frac{z_K}{\sqrt{\tilde{\Sigma}_{KK}}}\right|\right\} \\
&\qquad \leq \sqrt{\sum_{k=1}^{K} z_k^2 \left(\frac{1}{\sqrt{\Sigma_{kk}}} - \frac{1}{\sqrt{\tilde{\Sigma}_{kk}}}\right)^2} \\
&\qquad \leq \lambda_{\max}(\Sigma^{-1/2})\lambda_{\max}(\tilde{\Sigma}^{-1/2})\sqrt{\sum_{k=1}^{K} z_k^2 \left(\sqrt{\Sigma_{kk}} - \sqrt{\tilde{\Sigma}_{kk}}\right)^2} \\
&\qquad = \lambda_{\max}(\Sigma^{-1/2})\lambda_{\max}(\tilde{\Sigma}^{-1/2})\sqrt{\sum_{k=1}^{K} z_k^2 \left(\frac{\Sigma_{kk} - \tilde{\Sigma}_{kk}}{\sqrt{\Sigma_{kk}} + \sqrt{\tilde{\Sigma}_{kk}}}\right)^2} \\
&\qquad \leq \frac{1}{\lambda_{\min}(\Sigma^{1/2})} \frac{1}{\lambda_{\min}(\tilde{\Sigma}^{1/2})} \frac{1}{\lambda_{\min}(\Sigma^{1/2}) + \lambda_{\min}(\tilde{\Sigma}^{1/2})} \sqrt{\sum_{k=1}^{K} z_k^2 \left(\Sigma_{kk} - \tilde{\Sigma}_{kk}\right)^2} \\
&\qquad \leq \frac{1}{\lambda_{\min}(\Sigma^{1/2})} \frac{1}{\lambda_{\min}(\tilde{\Sigma}^{1/2})} \frac{1}{\lambda_{\min}(\Sigma^{1/2}) + \lambda_{\min}(\tilde{\Sigma}^{1/2})} \|\Sigma - \tilde{\Sigma}\|_S \|z\|.
\end{aligned}
$$

$\square$

Proof of Theorem 1.3. We prove the theorem by verifying the assumptions of Theorem 1.1 with $\kappa_n = \sqrt{nh_n}$. Before we do so notice that

$$\begin{aligned} P\left(\max_{k=1,\ldots,K_n} |\hat{f}_X(x_k) - f_X(x_k)| \geq \varepsilon\right) &\leq \sum_{k=1}^{K_n} P\left(|\hat{f}_X(x_k) - f_X(x_k)| \geq \varepsilon\right) \\ &\leq \sum_{k=1}^{K_n} \frac{E\left(\left(\hat{f}_X(x_k) - f_X(x_k)\right)^2\right)}{\varepsilon^2}. \end{aligned}$$

Moreover, with $A = \int K(u)u^2 du$ and $B = \int K(u)^2 du$ it is easy to show that

$$\left|E\left(\hat{f}_X(x_k)\right) - f_X(x_k)\right| \leq h_n^2 \frac{1}{2} A \sup_{x\in\mathcal{X}} |f_X''(x)|.$$

and

$$\left|Var\left(\hat{f}_X(x_k)\right)\right| \leq \frac{1}{nh_n} B \sup_{x\in\mathcal{X}} |f_X(x)| + \frac{1}{n}\left(f_X(x_k) + h_n^2 \frac{1}{2} A \sup_{x\in\mathcal{X}} |f_X''(x)|\right)^2.$$

It follows that for some constant C

$$P\left(\max_{k=1,\ldots,K_n} |\hat{f}_X(x_k) - f_X(x_k)| \geq \varepsilon\right) \leq \frac{C}{\varepsilon^2} K_n \left(h_n^4 + \frac{1}{nh_n}\right)$$

and thus

$$\max_{k=1,\ldots,K_n} |\hat{f}_X(x_k) - f_X(x_k)|^2 = O_p\left(K_n\left(h_n^4 + \frac{1}{nh_n}\right)\right)$$

uniformly over $P \in \mathcal{P}$. Similarly.

$$\begin{aligned} |\hat{\sigma}^2(x_k) - \sigma^2(x_k)| = &\left|\frac{\sum_{i=1}^n Y_i^2 K\left(\frac{x_k - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_k - X_i}{h_n}\right)} - E(Y^2 \mid X = x_k)\right| \\ &+ \left|\left(\frac{\sum_{i=1}^n Y_i K\left(\frac{x_k - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_k - X_i}{h_n}\right)}\right)^2 - (E(Y \mid X = x_k))^2\right| \end{aligned}$$

For the first term, let $h(x) = E(Y^2 \mid X = x)$ and let $V_i = Y_i^2 - h(X_i)$. Then

$$\frac{\sum_{i=1}^n Y_i^2 K\left(\frac{x_k - X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x_k - X_i}{h_n}\right)} - h(x_k) = \frac{\hat{m}_1(x_k)}{\hat{f}_X(x_k)} + \frac{\hat{m}_2(x_k)}{\hat{f}_X(x_k)},$$

where

$$\hat{m}_1(x_k) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x_k - X_i}{h_n}\right)(h(X_i) - h(x_k))$$

and

$$\hat{m}_2(x_k) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x_k - X_i}{h_n}\right)V_i.$$

Since $f_X(x)$ is bounded away from 0 and $\max_{k=1,\ldots,K_n}|\hat{f}_X(x_k) - f_X(x_k)|^2 = O_p(K_n(h_n^4 + 1/(nh_n)))$ it follows from similar arguments as above that

$$\max_{k=1,\ldots,K_n}\left|\frac{\sum_{i=1}^n Y_i^2 K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} - E(Y^2 \mid X = x_k)\right|^2 = O_p\left(K_n\left(h_n^4 + \frac{1}{nh_n}\right)\right)$$

uniformly over $P \in \mathcal{P}$. Similarly,

$$\max_{k=1,\ldots,K_n}\left|\frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} - E(Y \mid X = x_k)\right|^2 = O_p\left(K_n\left(h_n^4 + \frac{1}{nh_n}\right)\right)$$

and thus, with $\hat{E}(Y \mid X = x_k) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}$ have

$$\begin{aligned}
&\left|\left(\hat{E}(Y \mid X = x_k)\right)^2 - (E(Y \mid X = x_k))^2\right| \\
&\quad \leq \left|\hat{E}(Y \mid X = x_k) - E(Y \mid X = x_k)\right|\left|\hat{E}(Y \mid X = x_k) + E(Y \mid X = x_k)\right| \\
&\quad = O_p\left(\sqrt{K_n\left(h_n^4 + \frac{1}{nh_n}\right)}\right).
\end{aligned}$$

It follows that uniformly over $P \in \mathcal{P}$

$$\max_{k=1,\ldots,K_n}|\hat{\sigma}^2(x_k) - \sigma^2(x_k)|^2 = O_p\left(K_n\left(h_n^4 + \frac{1}{nh_n}\right)\right).$$

We can now verify the assumptions.

(1.1) Assumption 1.1 holds by construction with $\hat{\Omega} = \hat{\Sigma}^{-1}$ and $\|R_n\| = 0$.

(1.2) Let $G = (g_0(X_1), \ldots, g_0(X_n))'$, $Y = (Y_1, \ldots, Y_n)'$, and $U = (U_1, \ldots, U_n)'$. Now rewrite

$$\hat{\theta}_{ur,k} - \theta_{0,k} = \frac{\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x_k-X_i}{h_n}\right)(g_0(X_i) - g_0(x_k))}{\hat{f}_X(x_k)} + \frac{\frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x_k-X_i}{h_n}\right)U_i}{\hat{f}_X(x_k)}.$$

Arguments as above show that for some constant C

$$\left|\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x_k - X_i}{h_n}\right)(g_0(X_i) - g_0(x_k))\right| \leq Ch_n^2 + R_k$$

and $E(R_k^2) \leq C(h_n/n)$. Also notice that $\max_{k=1,\ldots,K_n} |\hat{f}_X(x_k) - f_X(x_k)| = o_p(1)$ and $f_X(x)$ is bounded away from 0.

We now apply a coupling argument to the $K_n \times 1$ vector with element

$$\sum_{i=1}^{n} K\Big(\frac{x_k - X_i}{h_n}\Big)U_i$$

to a normal random variable W. Let P_X be an $n \times K_n$ matrix with element (i,k) equal to $K((x_k - X_i)/h_n)$. Then, using the arguments as in the proof of Theorem 1.2, there exists

$$W \sim N\left(0, diag\left(\frac{1}{h_n}E\left(K((x_k - X_i)/h_n)^2\sigma^2(X_i)\right)\right)\right)$$

such that for any $\delta > 0$ and $\varepsilon_n > 0$

$$P\left(\left\|\frac{1}{\sqrt{nh_n}}P_X'U - W\right\| \geq 3\delta\varepsilon_n\right)$$
$$\leq C_0\frac{K_n^{3/2}\sum_{k=1}^{K_n} E\left(K\left(\frac{x_k - X_i}{h_n}\right)^3|U_i|^3\right)}{\sqrt{nh_n}\delta^3\varepsilon_n^3}\left(1 + \frac{\left|\log\left(\frac{K_n^{3/2}\sum_{k=1}^{K_n} E\left(K\left(\frac{x_k-X_i}{h_n}\right)^3|U_i|^3\right)}{\sqrt{nh_n}\delta^3\varepsilon_n^3}\right)\right|}{K_n}\right).$$

Since $E\left(K\left(\frac{x_k - X_i}{h_n}\right)^3|U_i|^3\right) \leq Mh_n$ for some constant $M < \infty$ we get for n large enough

$$P\left(\left\|\frac{1}{\sqrt{nh_n}}P_X'U - W\right\| \geq 3\delta\varepsilon_n\right) \leq C_0\frac{MK_n^{5/2}}{\sqrt{nh_n}\delta^3\varepsilon_n^3}\left(1 + \frac{\left|\log\left(\frac{MK_n^{5/2}}{\sqrt{nh_n}\delta^3\varepsilon_n^3}\right)\right|}{K_n}\right) = o(1).$$

Next let S_1 be a diagonal matrix with elements $\frac{1}{h_n}E\left(K((x_k - X_i)/h_n)^2\sigma^2(X_i)\right)$ and let S_2 be a diagonal matrix with elements $\sigma^2(x_k)f_X(x_k)B$. Our assumptions imply that for some constant C

$$\left|\frac{1}{h_n}E\left(K((x_k-X_i)/h_n)^2\sigma^2(X_i)\right)-\sigma^2(x_k)f_X(x_k)B\right|\le Ch_n.$$

Since also $\sigma^2(x_k)f_X(x_k)B$ is bounded away from 0 it follows that

$$\|W-S_2^{1/2}S_1^{-1/2}W\|\le O(h_n)\|W\|=O_p(h_n\sqrt{K_n}).$$

Let Q be a diagonal matrix with elements $f_X(x_k)$ and let $\hat{Q}$ be a diagonal matrix with elements $\frac{1}{nh_n}\sum_{i=1}^n K((x_k-X_i)/h_n)$. Then

$$\begin{aligned}
&\|\sqrt{nh_n}(\hat\theta_{ur}-\theta_0)-Q^{-1}S_2^{1/2}S_1^{-1/2}W\|\\
&\le\left\|\hat{Q}^{-1}\frac{1}{\sqrt{nh_n}}P_X'U-Q^{-1}S_2^{1/2}S_1^{-1/2}W\right\|+O_p(\sqrt{K_n}\sqrt{nh_n}h_n^2+\sqrt{K_n}h_n)\\
&\le\left\|\hat{Q}^{-1}\frac{1}{\sqrt{nh_n}}P_X'U-Q^{-1}W\right\|+O_p(\sqrt{K_n}\sqrt{nh_n}h_n^2+\sqrt{K_n}h_n)\\
&\le\left\|\hat{Q}^{-1}\frac{1}{\sqrt{nh_n}}P_X'U-\hat{Q}^{-1}W\right\|+\left\|(\hat{Q}^{-1}-Q)^{-1}W\right\|+O_p(\sqrt{K_n}\sqrt{nh_n}h_n^2+\sqrt{K_n}h_n)\\
&\le\lambda_{\max}(\hat{Q}^{-1})\left\|\frac{1}{\sqrt{nh_n}}P_X'U-W\right\|+\max_{k=1,\ldots,K_n}\left|\frac{1}{\hat{f}_X(x)}-\frac{1}{f_X(x)}\right|\|W\|\\
&\quad+O_p(\sqrt{K_n}\sqrt{nh_n}h_n^2+\sqrt{K_n}h_n)\\
&=O_p(1)\left\|\frac{1}{\sqrt{nh_n}}P_X'U-W\right\|+O_p\left(\sqrt{K_n^2\left(h_n^4+\frac{1}{nh_n}\right)}\right)\\
&\quad+O_p(\sqrt{K_n}\sqrt{nh_n}h_n^2+\sqrt{K_n}h_n)\\
&=o_p(\varepsilon_n)+O_p\left(\sqrt{K_nh_n^5n\frac{K_n}{nh_n}+\frac{K_n^5}{nh_n}\frac{1}{K_n^3}}\right)\\
&\quad+O_p(\sqrt{K_nnh_n^5}+(K_nh_n^5n)^{1/4}(K_n^5/(h_nn))^{1/4})\\
&=o_p(\varepsilon_n).
\end{aligned}$$

Now let $Z_n=Q^{-1}S_2^{1/2}S_1^{-1/2}W\sim N(0,\Sigma)$. It follows that for any $\delta>0$,

$$\sup_{P\in\mathcal{P}}P\left(\|\sqrt{n}(\hat\theta_{ur}-\theta_0)-Z_n\|\ge\delta\varepsilon_n\right)\to 0.$$

The result now follows because $\Omega=\Sigma^{-1}$ and $\lambda_{\max}(\Sigma)$ is uniformly bounded.

(1.3) Notice that $\lambda_{\max}(\Sigma)$ is bounded and bounded away from 0. Therefore, $\lambda_{\min}(\Omega)$ is bounded and bounded away from 0. Sufficient conditions for Assumption 1.3 are thus

$$\max_{k=1,\ldots,K_n}\left|\frac{\hat{\sigma}^2(x_k)}{\hat{f}_X(x_k)}-\frac{\sigma^2(x_k)}{f_X(x_k)}\right|=o_p(\varepsilon_n/\sqrt{K_n})$$

and

$$\max_{k=1,\ldots,K_n}\left|\frac{\hat{f}_X(x_k)}{\hat{\sigma}^2(x_k)}-\frac{f_X(x_k)}{\sigma^2(x_k)}\right|=o_p(\varepsilon_n^2/K_n).$$

Notice that

$$\begin{aligned}
&\max_{k=1,\ldots,K_n}\left|\frac{\hat{\sigma}^2(x_k)}{\hat{f}_X(x_k)}-\frac{\sigma^2(x_k)}{f_X(x_k)}\right| \\
&\leq \max_{k=1,\ldots,K_n}\frac{\sigma^2(x_k)|f_X(x_k)-\hat{f}_X(x_k)|+f_X(x_k)|\sigma^2(x_k)-\hat{\sigma}^2(x_k)|}{\hat{f}_X(x_k)f_X(x_k)} \\
&= O_p\left(\sqrt{K_n\left(h_n^4+\frac{1}{nh_n}\right)}\right) \\
&= o_p\left(\sqrt{\varepsilon_n^8/K_n^5+\varepsilon_n^6/K_n^4}\right) \\
&= o_p\left(\varepsilon_n^3/K_n^2\right)
\end{aligned}$$

uniformly over $P\in\mathcal{P}$ and analogously

$$\max_{k=1,\ldots,K_n}\left|\frac{\hat{f}_X(x_k)}{\hat{\sigma}^2(x_k)}-\frac{f_X(x_k)}{\sigma^2(x_k)}\right|=o_p\left(\varepsilon_n^3/K_n^2\right).$$

(1.4) This assumption holds by assumption.

(1.5) This step is analogous to that of the proof of Theorem 1.2.

$\square$

Proof of Theorem 1.4. We verify the assumptions of Theorem 1.1 with $\kappa_n=\sqrt{n}$. Before we do so we prove some preliminary results. Let P_X be the $n\times K_n$ matrix, where the ith row is

$p_{K_n}(X_i)'$. Let $Q = E(p_{K_n}(X)p_{K_n}(X)')$ and notice that

$$\begin{aligned}
\lambda_{\max}(Q) &= \max_{\|v\|=1} v'Qv \\
&= \max_{\|v\|=1} E((p_{K_n}(X)'v)^2) \\
&= \max_{\|v\|=1} \int (p_{K_n}(x)'v)^2 f_X(x)dx \\
&\leq \left(\sup_{x\in\mathcal{X}} f_X(x)\right) \max_{\|v\|=1} \int (p_{K_n}(x)'v)^2 dx \\
&= \sup_{x\in\mathcal{X}} f_X(x) \max_{\|v\|=1} \|v\|^2, \\
&= \sup_{x\in\mathcal{X}} f_X(x)
\end{aligned}$$

where the fifth line follows from the assumption that the basis functions are orthonormal. Similarly,

$$\lambda_{\min}(Q) \geq \inf_{x\in\mathcal{X}} f_X(x) > 0.$$

It then also follows that the maximum eigenvalue of Q^{-1} is bounded and the minimum eigenvalue of Q^{-1} is bounded away from 0. Since

$$\|Q^{-1}v\|^2 \leq \lambda_{\max}(Q^{-1}Q^{-1})\|v\|^2 = \lambda_{\max}(Q^{-1})^2\|v\|^2$$

we also have

$$\begin{aligned}
\lambda_{\max}(\Sigma) &= \max_{\|v\|=1} v'\Sigma v \\
&= \max_{\|v\|=1} (Q^{-1}v)'E(p_{K_n}(X)p_{K_n}(X)'U^2)(Q^{-1}v) \\
&= \max_{\|w\|=\lambda_{\max}(Q^{-1})} w'E(p_{K_n}(X)p_{K_n}(X)'U^2)w \\
&\leq \sup_{x\in\mathcal{X}} E(U^2 \mid X=x) \max_{\|w\|=\lambda_{\max}(Q^{-1})} w'E(p_{K_n}(X)p_{K_n}(X)')w \\
&= \sup_{x\in\mathcal{X}} E(U^2 \mid X=x) \sup_{x\in\mathcal{X}} f_X(x)\lambda_{\max}(Q^{-1})^2 \\
&< \infty
\end{aligned}$$

and

$$\lambda_{\min}(\Sigma) \geq \inf_{x\in\mathcal{X}} E(U^2 \mid X=x) \inf_{x\in\mathcal{X}} f_X(x)\lambda_{\min}(Q^{-1})^2 > 0.$$

Let $\hat{Q} = \frac{1}{n}(P_X' P_X)$. Then, similar as in Newey (1997),

$$\begin{aligned} E\|\hat{Q} - Q\|^2 &= E\sum_{j=1}^{K_n}\sum_{k=1}^{K_n}\left(\left(\hat{Q} - Q\right)_{jk}^2\right) \\ &\leq \frac{1}{n}\sum_{j=1}^{K_n}\sum_{k=1}^{K_n} E\left(p_j(X)^2 p_k(X)^2\right) \\ &\leq \xi(K_n)^2 \frac{1}{n} E\left(\sum_{k=1}^{K_n} p_k(X)^2\right) \\ &\leq \frac{\xi(K_n)^2 K_n}{n} \sup_{x\in\mathcal{X}} f_X(x) \end{aligned}$$

By Markov's inequality, it follows that

$$\|\hat{Q} - Q\| = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right) = o_p(1)$$

uniformly over $P \in \mathcal{P}$. Then by Lemma A.1

$$\left|\lambda_{\min}(\hat{Q}) - \lambda_{\min}(Q)\right| = o_p(1)$$

and

$$\left|\lambda_{\max}(\hat{Q}) - \lambda_{\max}(Q)\right| = o_p(1)$$

uniformly over $P \in \mathcal{P}$. It follows that $\lambda_{\min}(\hat{Q}) \geq \inf_{x\in\mathcal{X}} f_X(x)/2$ with probability approaching 1 and $\lambda_{\max}(\hat{Q})$ is bounded by $\sup_{x\in\mathcal{X}} f_X(x) + 1$ with probability approaching 1.

We now verify the assumptions.

(1.1) As discussed in Section 1.4.3, the assumption holds with $\|R_n\| = 0$ and $\hat{\Omega} = (1/n)(P_X' P_X)$ because the objective function is quadratic in θ.

(1.2) First let $G = (g_0(X_1), \ldots, g_0(X_n))'$, $Y = (Y_1, \ldots, Y_n)'$, and $U = (U_1, \ldots, U_n)'$. Now write

$$\hat{\theta}_{ur} = (P_X' P_X)^{-1} P_X' Y = (P_X' P_X)^{-1} P_X' G + (P_X' P_X)^{-1} P_X' U.$$

Let $Q = E(p_{K_n}(X)p_{K_n}(X)')$ and $\hat{Q} = \frac{1}{n}(P_X' P_X)$. We will apply the coupling argument to $\frac{1}{\sqrt{n}}P_X' U$. In particular, there exists $W \sim N(0, E(p_{K_n}(X)p_{K_n}(X)'U^2))$ such that for any $\delta > 0$ and $\varepsilon_n > 0$ and some $C_0 > 0$

$$P\left(\left\|\frac{1}{\sqrt{n}}P_X'U - W\right\| \geq 3\delta\varepsilon_n\right) \leq C_0\left(\frac{K_n}{n^{1/2}}\frac{E[\|p_{K_n}(X)U\|^3]}{\delta^3\varepsilon_n^3}\right)\left(1 + \frac{\left|\log\left(\frac{K_n}{n^{1/2}}\frac{E[\|p_{K_n}(X)U\|^3]}{\delta^3\varepsilon_n^3}\right)\right|}{K_n}\right).$$

Since $E(|U|^3 \mid X) \leq M$ for some $M < \infty$, we get

$$E[\|p_{K_n}(X)U\|^3] \leq ME[\|p_{K_n}(X)\|^3] \leq M\xi(K_n)E[\|p_{K_n}(X)\|^2] \leq O(\xi(K_n)K_n)$$

and it follows that uniformly over $P \in \mathcal{P}$

$$P\left(\left\|\frac{1}{\sqrt{n}}P_X'U - W\right\| \geq 3\delta\varepsilon_n\right) = O\left(\frac{K_n^2\xi(K_n)}{n^{1/2}\delta^3\varepsilon_n^3}\left|\log\left(\frac{K_n^2\xi(K_n)}{n^{1/2}\delta^3\varepsilon_n^3}\right)\right|\right) = o(1).$$

Next write

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{ur} - \theta_0) - Q^{-1}W &= \sqrt{n}(P_X'P_X)^{-1}P'(G - P_X\theta_0) + \hat{Q}^{-1}\frac{1}{\sqrt{n}}P'U - Q^{-1}W \\
&= \sqrt{n}(P_X'P_X)^{-1}P'(G - P_X\theta_0) + \left(\hat{Q}^{-1} - Q^{-1}\right)W \\
&\quad + \hat{Q}^{-1}\left(\frac{1}{\sqrt{n}}P'U - W\right).
\end{aligned}$$

Arguments in Newey (1997) imply that uniformly over $P \in \mathcal{P}$

$$\|\sqrt{n}(P_X'P_X)^{-1}P_X'(G - P_X\theta_0)\|^2 \leq \lambda_{\max}(\hat{Q}^{-1})O(nK_n^{-2\gamma}).$$

Since $\lambda_{\max}(\hat{Q}^{-1}) = O_p(1)$ we have

$$\|\sqrt{n}(P_X'P_X)^{-1}P_X'(G - P_X\theta_0)\|^2 \leq O_p(nK_n^{-2\gamma}).$$

Next write

$$\left(\hat{Q}^{-1} - Q^{-1}\right)W = Q^{-1}\left(Q - \hat{Q}\right)\hat{Q}^{-1}W.$$

Then

$$\left\|\left(\hat{Q}^{-1}-Q^{-1}\right)W\right\| \leq \lambda_{\max}(Q^{-1})\lambda_{\max}(\hat{Q}^{-1})\|Q-\hat{Q}\|\|W\|.$$

We also have

$$E(\|W\|^2)=\sum_{k=1}^{K_n}E(p_{K_n,k}(X)^2U^2)\leq K_n\sup_{x\in\mathcal{X}}f_X(x)\sup_{x\in\mathcal{X}}E(U^2\mid X=x).$$

By Markov's inequality $\|W\|=O_p(\sqrt{K_n})$ and therefore uniformly over $P\in\mathcal{P}$

$$\left\|\left(\hat{Q}^{-1}-Q^{-1}\right)W\right\|=O_p\left(\xi(K_n)\sqrt{\frac{K_n^2}{n}}\right).$$

Similarly,

$$\left\|\hat{Q}^{-1}\left(\frac{1}{\sqrt{n}}P_X'U-W\right)\right\|\leq O_p(1)\left\|\frac{1}{\sqrt{n}}P_X'U-W\right\|.$$

Putting these results together we get for any $\delta>0$

$$\begin{aligned}&P\left(\|\sqrt{n}(\hat{\theta}_{ur}-\theta_0)-Q^{-1}W\|\geq\delta\varepsilon_n\right)\\&\qquad\leq P\left(O_p(1)\left\|\frac{1}{\sqrt{n}}P_X'U-W\right\|+O_p\left(\sqrt{n}K_n^{-\gamma}+\frac{\xi(K_n)K_n}{\sqrt{n}}\right)\geq\delta\varepsilon_n\right)\end{aligned}$$

and the $O_p\left(\sqrt{n}K_n^{-\gamma}+\frac{\xi(K_n)K_n}{\sqrt{n}}\right)$ term does not depend on $P\in\mathcal{P}$. Therefore, for any ε_n satisfying the rate conditions

$$\begin{aligned}&\sup_{P\in\mathcal{P}}P\left(\|\sqrt{n}(\hat{\theta}_r-\theta_0)-Q^{-1}W\|\geq\delta\varepsilon_n\right)\\&\qquad\leq P\left(O_p(1)\left\|\frac{1}{\sqrt{n}}P_X'U-W\right\|\geq\frac{1}{2}\delta\varepsilon_n\right)+o(1).\end{aligned}$$

By the coupling argument for any $\delta>0$

$$\sup_{P\in\mathcal{P}}P\left(\|\sqrt{n}(\hat{\theta}_{ur}-\theta_0)-Q^{-1}W\|\geq\delta\varepsilon_n\right)\to 0$$

or

$$\sup_{P\in\mathcal{P}}P\left(\|\sqrt{n}(\hat{\theta}_{ur}-\theta_0)-Z_n\|\geq\delta\varepsilon_n\right)\to 0,$$

where $Z_n=Q^{-1}W\sim N(0,\Sigma)$. The result now follows because $\lambda_{\min}(\Omega)$ is uniformly bounded away from 0.

(1.3) As shown above $\lambda_{\min}(\Omega)$ is bounded from below and $\lambda_{\max}(\Sigma)$ is bounded from above. Moreover,

$$\|\hat{\Omega} - \Omega\|_S^2 \leq O_p\left(\xi(K_n)^2 \frac{K_n}{n}\right) = o_p(\varepsilon_n^4 / K_n^2)$$

uniformly over $P \in \mathcal{P}$.

Next define

$$\hat{D} = \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' \hat{U}_i^2$$

and $D = E(p_{K_n}(X_i) p_{K_n}(X_i)' U_i^2)$. Let $\hat{g}_{ur}(X_i) = p_{K_n}(X_i)' \hat{\theta}_{ur}$. Then

$$\begin{aligned}
\|\hat{D} - D\| &= \left\| \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' \hat{U}_i^2 - D \right\| \\
&\leq \left\| \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' U_i^2 - D \right\| \\
&\quad + \left\| \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' (g_0(X_i) - \hat{g}_{ur}(X_i))^2 \right\| \\
&\quad + 2 \left\| \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' U_i (g_0(X_i) - \hat{g}_{ur}(X_i)) \right\|.
\end{aligned}$$

Arguments as above show that

$$\left\| \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' U_i^2 - D \right\| = O_p\left(\xi(K_n) \sqrt{\frac{K_n}{n}}\right).$$

We have

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^n p_{K_n}(X_i) p_{K_n}(X_i)' (g_0(x_i) - \hat{g}_{ur}(X_i))^2 \right\| \\
&= \sqrt{\sum_{j=1}^{K_n} \sum_{k=1}^{K_n} \left(\frac{1}{n} \sum_{i=1}^n p_{K_n,j}(X_i) p_{K_n,k}(X_i) (g_0(X_i) - \hat{g}_{ur}(X_i))^2 \right)^2} \\
&\leq \sqrt{\sum_{j=1}^{K_n} \sum_{k=1}^{K_n} \left(\frac{1}{n} \sum_{i=1}^n p_{K_n,j}(X_i)^2 p_{K_n,k}(X_i)^2 \right)} \sup_{x \in \mathcal{X}} |g_0(X_i) - \hat{g}_{ur}(X_i)|^2.
\end{aligned}$$

Arguments as above show that

$$E\left(\sum_{j=1}^{K_n}\sum_{k=1}^{K_n}\left(\frac{1}{n}\sum_{i=1}^{n}p_{K_n,j}(X_i)^2p_{K_n,k}(X_i)^2\right)\right)=O\left(\xi(K_n)^2K_n\right).$$

Theorem 1 in Newey (1997) implies that

$$|g_0(x)-\hat{g}_{ur}(x)|^2=O_p\left(\xi(K_n)^2\frac{K_n}{n}+\xi(K_n)^2K_n^{-2\gamma}\right)=O_p\left(\xi(K_n)^2\frac{K_n}{n}\right)$$

and it is easy to show that the upper bound is uniform over $P\in\mathcal{P}$. Therefore

$$\sqrt{\sum_{j=1}^{K_n}\sum_{k=1}^{K_n}\left(\frac{1}{n}\sum_{i=1}^{n}p_{K_n,j}(X_i)^2p_{K_n,k}(X_i)^2\right)}\sup_{x\in\mathcal{X}}|g_0(X_i)-\hat{g}_{ur}(X_i)|^2=O_p\left(\xi(K_n)^3\frac{K_n^{3/2}}{n}\right)$$

uniformly over $P\in\mathcal{P}$. Next write

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=1}^{n}p_{K_n}(X_i)p_{K_n}(X_i)'U_i(g_0(X_i)-\hat{g}_{ur}(X_i))\right\|\\&=\sqrt{\sum_{j=1}^{K_n}\sum_{k=1}^{K_n}\left(\frac{1}{n}\sum_{i=1}^{n}p_{K_n,j}(X_i)p_{K_n,k}(X_i)U_i(g_0(X_i)-\hat{g}_{ur}(X_i))\right)^2}\\&\leq\sqrt{\sum_{j=1}^{K_n}\sum_{k=1}^{K_n}\left(\frac{1}{n}\sum_{i=1}^{n}p_{K_n,j}(X_i)^2p_{K_n,k}(X_i)^2U_i^2\right)}\sup_{x\in\mathcal{X}}|g_0(x)-\hat{g}_{ur}(x)|.\end{aligned}$$

Moreover analogous arguments as before yield

$$\begin{aligned}&\sqrt{\sum_{j=1}^{K_n}\sum_{k=1}^{K_n}\left(\frac{1}{n}\sum_{i=1}^{n}p_{K_n,j}(X_i)^2p_{K_n,k}(X_i)^2U_i^2\right)}\sup_{x\in\mathcal{X}}|g_0(x)-\hat{g}_{ur}(x)|\\&\qquad=O_p\left(\xi(K_n)^2\tfrac{K_n}{\sqrt{n}}\right).\end{aligned}$$

It follows that

$$\|\hat{D}-D\|=O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right)+O_p\left(\xi(K_n)^3\frac{K_n^{3/2}}{n}\right)+O_p\left(\xi(K_n)^2\frac{K_n}{\sqrt{n}}\right).$$

Next for any $v \in \mathbb{R}^{K_n}$ such that $\|v\| = 1$ we get

$$\begin{aligned}\left\|\hat{\Sigma}v - \Sigma v\right\| &= \|\hat{Q}^{-1}\hat{D}\hat{Q}^{-1}v - Q^{-1}DQ^{-1}v\| \\ &= \|Q^{-1}\hat{D}\hat{Q}^{-1}v - Q^{-1}DQ^{-1}v + (\hat{Q}^{-1} - Q^{-1})\hat{D}\hat{Q}^{-1}v\| \\ &\leq \lambda_{\min}(Q^{-1})\|\hat{D}\hat{Q}^{-1}v - DQ^{-1}v\| \\ &\quad + \sqrt{\lambda_{\max}((\hat{Q}^{-1} - Q^{-1})(\hat{Q}^{-1} - Q^{-1}))}\lambda_{\max}(\hat{D})\lambda_{\max}(\hat{Q}^{-1}).\end{aligned}$$

We also have

$$\|(\hat{Q}^{-1} - Q^{-1})v\| = \|Q^{-1}(Q - \hat{Q})\hat{Q}^{-1}v\| \leq \lambda_{\max}(Q^{-1})\lambda_{\max}(\hat{Q}^{-1})\|Q - \hat{Q}\|$$

Thus,

$$\sqrt{\lambda_{\max}((\hat{Q}^{-1} - Q^{-1})(\hat{Q}^{-1} - Q^{-1}))} = O_p(1)\|Q - \hat{Q}\|.$$

It is also easy to show that the maximum and minimum eigenvalues of D are uniformly bounded and bounded away from 0 and thus, the same is true for Σ. Therefore, using arguments as above

$$\begin{aligned}\|\hat{\Sigma} - \Sigma\|_S &= O_p(1)(\|Q - \hat{Q}\| + \|D - \hat{D}\|) \\ &= O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right) + O_p\left(\xi(K_n)^3\frac{K_n^{3/2}}{n}\right) + O_p\left(\xi(K_n)^2\frac{K_n}{\sqrt{n}}\right).\end{aligned}$$

Consequently, Assumption 1.3 holds.

(1.4) The assumption holds by assumption.

(1.5) Let $\|z_1\|_T = \sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'z_1}{\sigma(x)}\right|$, which is a norm since $E(p_{K_n}(X)p_{K_n}(X)')$ has full rank. By the reverse triangle inequality

$$|T(z_1, \Sigma) - T(z_2, \Sigma)| \leq \sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'(z_1 - z_2)}{\sigma(x)}\right| \leq \sup_{x\in\mathcal{X}}\frac{\|p_{K_n}(x)\|}{\sigma(x)}\|z_1 - z_2\|.$$

But

$$\left(\sup_{x\in\mathcal{X}}\frac{\|p_{K_n}(x)\|}{\sigma(x)}\right)^2 \leq \sup_{p\in\mathbb{R}^{K_n}:\,\|p\|=1}\frac{p'p}{p'\Sigma p} = \frac{1}{\lambda_{\min}(\Sigma)} = \lambda_{\max}(\Sigma^{-1}).$$

Moreover, let $\sigma_1(x) = \sqrt{p_{K_n}(x)'\Sigma_1 p_{K_n}(x)}$ and $\sigma_2(x) = \sqrt{p_{K_n}(x)'\Sigma_2 p_{K_n}(x)}$. Then

$$\begin{aligned}
|T(z,\Sigma_1) - T(z,\Sigma_2)| &= \left|\sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'z}{\sigma_1(x)}\right| - \sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'z}{\sigma_2(x)}\right|\right| \\
&\leq \sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'z}{\sigma_1(x)} - \frac{p_{K_n}(x)'z}{\sigma_2(x)}\right| \\
&= \sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'z}{\sigma_1(x)}\left(1 - \frac{\sigma_1(x)}{\sigma_2(x)}\right)\right| \\
&\leq \frac{1}{\sqrt{\lambda_{\min}(\Sigma_1)}}\|z\| \sup_{x\in\mathcal{X}}\left|\left(1 - \frac{\sigma_1(x)}{\sigma_2(x)}\right)\right|.
\end{aligned}$$

Finally

$$\begin{aligned}
\sup_{x\in\mathcal{X}}\left|\left(1 - \frac{\sigma_1(x)}{\sigma_2(x)}\right)\right| &= \sup_{x\in\mathcal{X}}\left|\frac{\sigma_1(x) - \sigma_2(x)}{\sigma_2(x)}\right| \\
&= \sup_{x\in\mathcal{X}}\left|\frac{\frac{\sqrt{p_{K_n}(x)'\Sigma_1 p_{K_n}(x)}}{\|p_{K_n}(x)\|} - \frac{\sqrt{p_{K_n}(x)'\Sigma_2 p_{K_n}(x)}}{\|p_{K_n}(x)\|}}{\frac{\sqrt{p_{K_n}(x)'\Sigma_2 p_{K_n}(x)}}{\|p_{K_n}(x)\|}}\right| \\
&\leq \frac{1}{\sqrt{\lambda_{\min}(\Sigma_2)}}\sup_{\|v\|=1}\left|\sqrt{v'\Sigma_1 v} - \sqrt{v'\Sigma_2 v}\right| \\
&= \frac{1}{\sqrt{\lambda_{\min}(\Sigma_2)}}\sup_{\|v\|=1}\left|\frac{v'\Sigma_1 v - v'\Sigma_2 v}{\sqrt{v'\Sigma_1 v} + \sqrt{v'\Sigma_2 v}}\right| \\
&\leq \frac{1}{\sqrt{\lambda_{\min}(\Sigma_2)}}\frac{1}{\sqrt{\lambda_{\min}(\Sigma_1)} + \sqrt{\lambda_{\min}(\Sigma_2)}}\|\Sigma_1 - \Sigma_2\|_S.
\end{aligned}$$

□

Proof of Corollary 1.2. Let

$$a_n(x) = \frac{\sqrt{n}(g_0(x) - p_{K_n}(x)'\theta_0)}{\hat{\sigma}(x)}$$

and notice

$$|a_n(x)| \leq \frac{p_{K_n}(x)'p_{K_n}(x)}{p_{K_n}(x)'\hat{\Sigma}p_{K_n}(x)}\frac{1}{\|p_{K_n}(x)\|^2}C_g\sqrt{n}K_n^{-\gamma} \leq \frac{1}{\lambda_{\min}(\hat{\Sigma})}\frac{1}{\inf_{x\in\mathcal{X}}\|p_{K_n}(x)\|^2}C_g\sqrt{n}K_n^{-\gamma}.$$

Hence for any $\delta > 0$

$$\sup_{P\in\mathcal{P}} P\left(\sup_{x\in\mathcal{X}}|a_n(x)| \geq \delta\varepsilon_n\right) = o(1).$$

Next notice that if

$$\sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r-\theta_0)\right)}{\hat{\sigma}(x)}\right| \leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})-\sup_{x\in\mathcal{X}}|a(x)|,$$

then for all $x \in \mathcal{X}$

$$-\hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})+\hat{\sigma}(x)\sup_{x\in\mathcal{X}}|a(x)| \leq p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r-\theta_0)\right)$$

and

$$p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r-\theta_0)\right) \leq \hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})-\hat{\sigma}(x)\sup_{x\in\mathcal{X}}|a(x)|,$$

which implies by the definition of $a_n(x)$ that for all $x \in \mathcal{X}$

$$p_{K_n}(x)'\hat{\theta}_r-\frac{\hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})}{\sqrt{n}} \leq g_0(x) \leq p_{K_n}(x)'\hat{\theta}_r+\frac{\hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})}{\sqrt{n}}.$$

Finally, if

$$\sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r-\theta_0)\right)}{\hat{\sigma}(x)}\right| \leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})-\sup_{x\in\mathcal{X}}|a(x)|,$$

then $\theta_0 \in CI$, and therefore for all $x \in \mathcal{X}$

$$\hat{g}_l(x) \leq p_{K_n}(x)'\hat{\theta}_r-\frac{\hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})}{\sqrt{n}} \text{ and } \hat{g}_u(x) \geq p_{K_n}(x)'\hat{\theta}_r+\frac{\hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})}{\sqrt{n}}.$$

We conclude that

$$\begin{aligned}
P&\left(\sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r-\theta_0)\right)}{\hat{\sigma}(x)}\right| \leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})-\sup_{x\in\mathcal{X}}|a(x)|\right)\\
&\leq P\left(\left|p_{K_n}(x)'\hat{\theta}_r-g_0(x)\right| \leq \frac{\hat{\sigma}(x)c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})}{\sqrt{n}}\ \forall x\in\mathcal{X}\right)\\
&\leq P\left(\hat{g}_l(x) \leq g_0(x) \leq \hat{g}_u(x)\ \ \forall x \in \mathcal{X}\right).
\end{aligned}$$

The proof of Theorem 1.1 implies that under Assumption 1.1 - 1.6 for any δ small enough

$$\sup_{P\in\mathcal{P}}\left|P\left(T(\sqrt{n}(\hat{\theta}_r-\theta_0),\hat{\Sigma}) \leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma},\hat{\Omega})-\delta\varepsilon_n\right)-(1-\alpha)\right| \to 0.$$

Since for any $\delta > 0$

$$\sup_{P\in\mathcal{P}} P\left(\sup_{x\in\mathcal{X}} |a_n(x)| \geq \delta\varepsilon_n\right) = o(1),$$

Theorem 1.4 implies that

$$\sup_{P\in\mathcal{P}}\left|P\left(\sup_{x\in\mathcal{X}}\left|\frac{p_{K_n}(x)'\left(\sqrt{n}(\hat{\theta}_r-\theta_0)\right)}{\hat{\sigma}(x)}\right| \leq c_{1-\alpha,n}(\theta_0,\hat{\Sigma}) - \sup_{x\in\mathcal{X}}|a(x)|\right) - (1-\alpha)\right| \to 0.$$

□

Proof of Theorem 1.5. We verify the Assumptions of Theorem 1.1 with $\kappa_n = \sqrt{n}$. Before we do so, let

$$\hat{Q}_{XZ} = \frac{1}{n}\sum_{i=1}^{n} p_{K_n}(X_i)p_{K_n}(Z_i) \text{ and } Q_{XZ} = E\left(p_{K_n}(X)p_{K_n}(Z)'\right)$$

and

$$\hat{Q}_{Z} = \frac{1}{n}\sum_{i=1}^{n} p_{K_n}(Z_i)p_{K_n}(Z_i)' \text{ and } Q_{Z} = E\left(p_{K_n}(Z)p_{K_n}(Z)'\right).$$

Arguments as in the proof of Theorem 1.4 show that

$$\|\hat{Q}_{XZ} - Q_{XZ}\| = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right)$$

and

$$\left\|\hat{Q}_{Z} - Q_{Z}\right\| = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right)$$

uniformly over $P \in \mathcal{P}$. We also have

$$\left|\|\hat{Q}_{XZ}\|_S - \|Q_{XZ}\|_S\right| \leq \left\|\hat{Q}_{XZ} - Q_{XZ}\right\|_S \leq \left\|\hat{Q}_{XZ} - Q_{XZ}\right\|,$$

which implies that

$$\left|\lambda_{\max}(Q_{XZ}Q_{XZ}') - \lambda_{\max}(\hat{Q}_{XZ}\hat{Q}_{XZ}')\right| = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right) = o_p(1).$$

Similarly,

$$\left|\lambda_{\max}(Q_Z) - \lambda_{\max}(\hat{Q}_Z)\right| = o_p(1).$$

It then also follows that uniformly over $P \in \mathcal{P}$

$$\|\hat{Q}_{XZ}\hat{Q}'_{XZ} - Q_{XZ}Q'_{XZ}\|_S = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right).$$

Moreover

$$\begin{aligned}\left|\sqrt{\lambda_{\min}(Q_{XZ}Q'_{XZ})} - \sqrt{\lambda_{\min}(\hat{Q}_{XZ}\hat{Q}'_{XZ})}\right| &= \left|\min_{\|v\|=1}\|Q_{XZ}v\| - \min_{\|v\|=1}\|\hat{Q}_{XZ}v\|\right| \\ &\leq \max_{\|v\|=1}\|(Q_{XZ} - \hat{Q}_{XZ})v\| \\ &\leq \|Q_{XZ} - \hat{Q}_{XZ}\|\end{aligned}$$

and thus

$$\frac{\left|\sqrt{\lambda_{\min}(Q_{XZ}Q'_{XZ})} - \sqrt{\lambda_{\min}(\hat{Q}_{XZ}\hat{Q}'_{XZ})}\right|}{\sqrt{\lambda_{\min}(Q_{XZ}Q'_{XZ})}} = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n\tau_{K_n}}}\right) = o_p(1)$$

and

$$\left|\frac{\lambda_{\min}(Q_{XZ}Q'_{XZ})}{\lambda_{\min}(\hat{Q}_{XZ}\hat{Q}'_{XZ})} - 1\right| \xrightarrow{p} 0.$$

(1.1) This assumption again holds with $\|R_n\| = 0$.

(1.2) First let $G = (g_0(X_1), \ldots, g_0(X_n))'$, $Y = (Y_1, \ldots, Y_n)'$, and $U = (U_1, \ldots, U_n)'$. Now write

$$\hat{\theta}_{ur} = (P'_Z P_X)^{-1} P'_Z Y = (P'_Z P_X)^{-1} P'_Z G + (P'_Z P_X)^{-1} P'_Z U.$$

Similar as before, there exists $W \sim N(0, \sigma^2 E(p_{K_n}(Z_i) p_{K_n}(Z_i)'))$ such that for $\delta > 0$ and $\varepsilon_n > 0$

$$\begin{aligned}&P\left(\left\|\frac{1}{\sqrt{n}}P'_Z U - W\right\| \geq 3\delta\varepsilon_n\right) \\ &\qquad \leq C_0\left(\frac{K_n}{n^{1/2}}\frac{E[\|p_{K_n}(Z_i)U_i\|^3]}{\delta^3\varepsilon_n^3}\right)\left(1 + \frac{\left|\log\left(\frac{K_n}{n^{1/2}}\frac{E[\|p_{K_n}(Z_i)U_i\|^3]}{\delta^3\varepsilon_n^3}\right)\right|}{K_n}\right).\end{aligned}$$

Since $E(|U_i|^3 \mid Z_i) \leq M$ for some $M < \infty$, we get

$$E[\|p_{K_n}(Z_i)U_i\|^3] \leq ME[\|p_{K_n}(Z_i)\|^3] \leq M\xi(K_n)E[\|p_{K_n}(Z_i)\|^2] \leq O(\xi(K_n)K_n)$$

and it follows that

$$P\left(\left\|\frac{1}{\sqrt{n}}P'U - W\right\| \geq 3\delta\varepsilon_n\right) = O\left(\frac{K_n^2\xi(K_n)}{n^{1/2}\delta^3\varepsilon_n^3}\left|\log\left(\frac{K_n^2\xi(K_n)}{n^{1/2}\delta^3\varepsilon_n^3}\right)\right|\right)$$

uniformly over $P \in \mathcal{P}$.

Next write

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_{ur} - \theta_0) - Q_{XZ}^{-1}W \\
&\qquad = \sqrt{n}(P_Z'P_X)^{-1}P_Z'(G - P_X\theta_0) + \hat{Q}_{XZ}^{-1}\frac{1}{\sqrt{n}}P_Z'U - Q_{XZ}^{-1}W \\
&\qquad = \sqrt{n}(P_Z'P_X)^{-1}P_Z'(G - P_X\theta_0) + \left(\hat{Q}_{XZ}^{-1} - Q_{XZ}^{-1}\right)W \\
&\qquad\quad + \hat{Q}_{XZ}^{-1}\left(\frac{1}{\sqrt{n}}P_Z'U - W\right).
\end{aligned}$$

We have

$$\begin{aligned}
&\|\sqrt{n}(P_Z'P_X)^{-1}P_Z'(G - P_X\theta_0)\|^2 \\
&= \frac{1}{n}(G - P_X\theta_0)'P_Z(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}P_Z'(G - P_X\theta_0) \\
&= \frac{1}{n}(G - P_X\theta_0)'P_Z Q_Z^{-1/2}Q_Z^{1/2}(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}Q_Z^{1/2}Q_Z^{-1/2}P_Z'(G - P_X\theta_0) \\
&\leq \lambda_{\max}(Q_Z^{1/2}(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}Q_Z^{1/2})\frac{1}{n}(G - P_X\theta_0)'P_Z Q_Z^{-1}P_Z'(G - P_X\theta_0) \\
&\leq \lambda_{\max}(Q_Z^{1/2}(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}Q_Z^{1/2})(G - P_X\theta_0)'P_Z(P_Z'P_Z)^{-1}P_Z'(G - P_X\theta_0) \\
&\leq \lambda_{\max}(Q_Z^{1/2}(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}Q_Z^{1/2})(G - P_X\theta_0)'(G - P_X\theta_0) \\
&\leq \lambda_{\max}(Q_Z^{1/2}(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}Q_Z^{1/2})O(nb(K_n)^2),
\end{aligned}$$

where the sixth line follows because $P_Z(P_Z'P_Z)^{-1}P_Z'$ is idempotent. Finally,

$$\begin{aligned}
\|(\hat{Q}_{XZ})^{-1}Q_Z^{1/2}v\|^2 &\leq \lambda_{\max}((\hat{Q}_{XZ})^{-1}(\hat{Q}_{XZ}')^{-1})\lambda_{\max}(Q_Z)\|v\|^2 \\
&= \frac{\lambda_{\max}(Q_Z)}{\lambda_{\min}(\hat{Q}_{XZ}\hat{Q}_{XZ}')}\|v\|^2
\end{aligned}$$

which implies that $\lambda_{\max}(Q_Z^{1/2}(\hat{Q}_{XZ}')^{-1}(\hat{Q}_{XZ})^{-1}Q_Z^{1/2}) = O_p\left(1/\tau_{K_n}\right)$ and

$$\|\sqrt{n}(P_Z'P_X)^{-1}P_Z'(G - P\theta_0)\|^2 = O_p(nb(K_n)^2/\tau_{K_n})$$

uniformly over $P \in \mathcal{P}$.

Next write

$$\left(\hat{Q}_{XZ}^{-1} - Q_{XZ}^{-1}\right) W = Q_{XZ}^{-1} \left(Q_{XZ} - \hat{Q}_{XZ}\right) \hat{Q}_{XZ}^{-1} W.$$

Then

$$\begin{aligned}
&\left\| \left(\hat{Q}_{XZ}^{-1} - Q_{XZ}^{-1}\right) W \right\| \\
&\qquad \leq \sqrt{\lambda_{\max}(Q_{XZ}^{-1}(Q'_{XZ})^{-1}) \lambda_{\max}(\hat{Q}_{XZ}^{-1}(\hat{Q}'_{XZ})^{-1})} \|Q_{XZ} - \hat{Q}_{XZ}\| \|W\|.
\end{aligned}$$

We also have

$$E(\|W\|^2) = \sigma^2 \sum_{k=1}^{K_n} E(p_{K_n,k}(Z_i)^2) \leq K_n \sup_{z \in \mathcal{Z}} f_Z(z) \sigma^2.$$

By Markov's inequality $\|W\| = O_p(\sqrt{K_n})$ and therefore

$$\left\| \left(\hat{Q}_{XZ}^{-1} - Q_{XZ}^{-1}\right) W \right\| = O_p\left(\xi(K_n) \sqrt{\frac{K_n^2}{n \tau_{K_n}^2}}\right).$$

Hence

$$\begin{aligned}
&\|\sqrt{n}(\hat{\theta}_{ur} - \theta_0) - Q_{XZ}^{-1} W\| \\
&\qquad \leq \left\| \hat{Q}_{XZ}^{-1} \left(\frac{1}{\sqrt{n}} P'_Z U - W\right) \right\| + O_p\left(\frac{\sqrt{n} b(K_n)}{\sqrt{\tau_{K_n}}} + \xi(K_n) \sqrt{\frac{K_n^2}{n \tau_{K_n}^2}}\right) \\
&\qquad \leq O_p\left(\left(\frac{K_n^4 \xi(K_n)^2}{n \tau_{K_n}^3}\right)^{1/6}\right) + O_p\left(\frac{\sqrt{n} b(K_n)}{\sqrt{\tau_{K_n}}} + \xi(K_n) \sqrt{\frac{K_n^2}{n \tau_{K_n}^2}}\right).
\end{aligned}$$

We have

$$\min_{\|v\|=1} v' \Omega v = \min_{\|v\|=1} \|Q_Z^{-1/2} Q'_{XZ} v\|^2 \geq \lambda_{\min}(Q_Z^{-1}) \min_{\|v\|=1} \|Q'_{XZ} v\|^2 = \lambda_{\min}(Q_Z^{-1}) \tau_{K_n}$$

and thus

$$\begin{aligned}
&\lambda_{\min}(\Omega)^{-1/2} \|\sqrt{n}(\hat{\theta}_{ur} - \theta_0) - Q_{XZ}^{-1} W\| \\
&\qquad \leq O_p\left(\left(\frac{K_n^4 \xi(K_n)^2}{n \tau_{K_n}^6}\right)^{1/6} + \frac{\sqrt{n} b(K_n)}{\tau_{K_n}} + \xi(K_n) \sqrt{\frac{K_n^2}{n \tau_{K_n}^3}}\right) \\
&\qquad = o_p(\varepsilon_n).
\end{aligned}$$

(1.3) We have

$$\hat{\Omega} = \overline{P}_{XZ} \left(\frac{1}{n} \sum_{i=1}^{n} p_{K_n}(Z_i) p_{K_n}(Z_i)' \right)^{-1} \overline{P}'_{XZ}$$

where

$$\overline{P}_{XZ} = \frac{1}{n} \sum_{i=1}^{n} p_{K_n}(X_i) p_{K_n}(Z_i)'$$

and

$$\Omega = E\left(p_{K_n}(X) p_{K_n}(Z)'\right) E\left(p_{K_n}(Z) p_{K_n}(Z)'\right)^{-1} E\left(p_{K_n}(Z) p_{K_n}(X)'\right).$$

It is easy to show using arguments as in the proof of Theorem 1.4 that $\lambda_{\max}(Q_Z)$ is bounded and $\lambda_{\min}(Q_Z)$ is bounded away from 0. We have

$$\begin{aligned}\max_{\|v\|=1} v'\Omega v = \max_{\|v\|=1} \|Q_Z^{-1/2} Q'_{XZ} v\|^2 &\geq \lambda_{\min}(Q_Z^{-1}) \max_{\|v\|=1} \|Q'_{XZ} v\|^2 \\ &= \lambda_{\min}(Q_Z^{-1}) \lambda_{\max}(Q_{XZ} Q'_{XZ}).\end{aligned}$$

which implies that $\lambda_{\max}(\Omega)$ is uniformly bounded from below. Similarly,

$$\lambda_{\max}(\Omega) \leq \lambda_{\max}(Q_{XZ} Q'_{XZ}) \lambda_{\max}(Q_Z^{-1}),$$

which implies that $\lambda_{\max}(\Omega)$ is bounded. Since

$$\lambda_{\max}(\Omega) = \frac{1}{\lambda_{\min}(\Omega^{-1})} = \frac{\sigma^2}{\lambda_{\min}(\Sigma)}$$

also $\lambda_{\min}(\Sigma)$ is bounded and bounded away from zero. Moreover, it follows that for all

v with $\|v\| = 1$

$$\begin{aligned}
&\|(\hat\Omega - \Omega)v\| \\
&\quad = \|Q_{XZ}Q_ZQ'_{XZ}v - \hat Q_{XZ}\hat Q_Z\hat Q'_{XZ}v\| \\
&\quad \le \|\hat Q_{XZ}(Q_ZQ'_{XZ} - \hat Q_Z\hat Q'_{XZ})v\| + \|(\hat Q_{XZ} - Q_{XZ})Q_ZQ'_{XZ}v\| \\
&\quad \le \sqrt{\lambda_{\max}(\hat Q_{XZ}\hat Q'_{XZ})}\|(Q_ZQ'_{XZ} - \hat Q_Z\hat Q'_{XZ})v\| \\
&\qquad\qquad + \sqrt{\lambda_{\max}(Q_{XZ}Q'_{XZ})}\lambda_{\max}(Q_Z)\|\hat Q_{XZ} - Q_{XZ}\| \\
&\quad \le \sqrt{\lambda_{\max}(\hat Q_{XZ}\hat Q'_{XZ})}\left(\sqrt{\lambda_{\max}(Q_{XZ}Q'_{XZ})}\|\hat Q_Z - Q_Z\| + \lambda_{\max}(\hat Q_Z)\|\hat Q_{XZ} - Q_{XZ}\|\right) \\
&\qquad + \sqrt{\lambda_{\max}(Q_{XZ}Q'_{XZ})}\lambda_{\max}(Q_Z)\|\hat Q_{XZ} - Q_{XZ}\|.
\end{aligned}$$

Thus

$$\|\hat\Omega - \Omega\|_S = O_p\left(\xi(K_n)\sqrt{\frac{K_n}{n}}\right).$$

It follows that

$$\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)^2}\|\hat\Omega - \Omega\|_S = O_p\left(\sqrt{\frac{\xi(K_n)^2K_n}{n\tau^6_{K_n}}}\right) = o_p(\varepsilon^2/K_n)$$

uniformly over $P \in \mathcal{P}$.

Next notice that

$$\begin{aligned}
\frac{\sqrt{\lambda_{\max}(\Sigma)}}{\sqrt{\lambda_{\min}(\Omega)}}\|\hat\Omega^{-1} - \Omega^{-1}\|_S &\le \frac{\sqrt{\lambda_{\max}(\Sigma)}\lambda_{\max}(\Omega^{-1})\lambda_{\max}(\hat\Omega^{-1})}{\sqrt{\lambda_{\min}(\Omega)}}\|\hat\Omega - \Omega\| \\
&= O_p\left(\sqrt{\frac{\xi(K_n)^2K_n}{n\tau^6_{K_n}}}\right)
\end{aligned}$$

uniformly over $P \in \mathcal{P}$. Therefore

$$\begin{aligned}
\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\hat\Sigma - \Sigma\|_S^2 &= \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\hat\sigma^2\hat\Omega^{-1} - \sigma^2\Omega^{-1}\|_S^2 \\
&= \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\hat\sigma^2(\hat\Omega^{-1} - \Omega^{-1}) + (\hat\sigma^2 - \sigma^2)\Omega^{-1}\|_S^2 \\
&\le 2\hat\sigma^2\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\hat\Omega^{-1} - \Omega^{-1}\|_S^2 + 2(\hat\sigma^2 - \sigma^2)^2\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)}\|\Omega^{-1}\|_S^2 \\
&\le o_p(\varepsilon_n^2/K_n) + 2(\hat\sigma^2 - \sigma^2)^2\frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Omega)^3}.
\end{aligned}$$

Hence, we have to show that $(\hat{\sigma}^2 - \sigma^2)^2 = o_p(\tau_{K_n}^4 \varepsilon_n^2 / K_n)$ uniformly over $P \in \mathcal{P}$. Write

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - p_{K_n}(X_i)'\hat{\theta}_{ur})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} U_i^2 + \frac{1}{n}\sum_{i=1}^{n}(p_{K_n}(X_i)'\hat{\theta}_{ur} - g_0(X_i))^2 \\
&\quad + \frac{2}{n}\sum_{i=1}^{n} U_i(p_{K_n}(X_i)'\hat{\theta}_{ur} - g_0(X_i)).
\end{aligned}$$

Since $\frac{1}{n}\sum_{i=1}^{n} U_i^2 - \sigma^2 = O_p(1/\sqrt{n})$ uniformly over $P \in \mathcal{P}$ it follows that

$$\begin{aligned}
\left| \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2 - \sigma^2 \right| &\leq O_p\left(\frac{1}{\sqrt{n}}\right) + 2\lambda_{\max}(\hat{Q}_X)\|\hat{\theta}_{ur} - \theta_0\|^2 + 2b(K_n)^2 \\
&\quad + \sqrt{\frac{1}{n}\sum_{i=1}^{n} U_i^2}\sqrt{2\lambda_{\max}(\hat{Q}_X)\|\hat{\theta}_{ur} - \theta_0\|^2 + 2b(K_n)^2}.
\end{aligned}$$

Moreover, the arguments above (verification of Assumption 1.2) imply that

$$\|\hat{\theta}_{ur} - \theta_0\|^2 = O_p\left(\frac{b(K_n)^2}{\tau_{K_n}} + \frac{K_n}{n\tau_{K_n}}\right).$$

Hence,

$$(\hat{\sigma}^2 - \sigma^2)^2 = O_p\left(\frac{b(K_n)^2}{\tau_{K_n}} + \frac{K_n}{n\tau_{K_n}}\right) = o_p(\tau_{K_n}^4 \varepsilon_n^2 / K_n)$$

uniformly over $P \in \mathcal{P}$.

(1.4) This assumption holds by assumption.

(1.5) This is identical to the arguments in the proof of Theorem 1.4.

□

A.3 Worst-case bias

In this section we briefly describe how we can incorporate a worst-case bias as in Armstrong and Kolesár (2016) instead of using the undersmoothing assumption. We focus on the kernel regression framework in Section 1.4.2, but the approach is also applicable in other settings. Let $\tilde{\theta}_0 = E(\hat{\theta}_{ur})$ and suppose that there exists a known constant C_B such that $\|\sqrt{nh_n}(\tilde{\theta}_0 - \theta_0)\| \leq C_B$. Now notice that under the assumptions of Theorem 1.3, but without the undersmoothing assumption, it holds that

$$\sup_{P \in \mathcal{P}} \left| P\left(T\left(\sqrt{nh_n}(\hat{\theta}_r - \tilde{\theta}_0), \hat{\Sigma}\right) \leq c_{1-\alpha,n}\left(\tilde{\theta}_0, \hat{\Sigma}, \hat{\Omega}\right)\right) - (1-\alpha)\right| \to 0.$$

Furthermore, from the arguments of the proof of Theorem 1.3 it follows that

$$\left|T\left(\sqrt{nh_n}(\hat{\theta}_r - \tilde{\theta}_0), \hat{\Sigma}\right) - T\left(\sqrt{nh_n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}\right)\right| \leq \frac{1}{\sqrt{\lambda_{\min}(\hat{\Sigma})}} \|\sqrt{nh_n}(\tilde{\theta}_0 - \theta_0)\|$$

and using similar arguments it is easy to show that

$$c_{1-\alpha,n}\left(\theta_0, \hat{\Sigma}, \hat{\Omega}\right) \leq c_{1-\alpha,n}\left(\tilde{\theta}_0, \hat{\Sigma}, \hat{\Omega}\right) + \frac{1}{\sqrt{\lambda_{\min}(\hat{\Sigma})}}\left(\sqrt{\frac{\lambda_{\max}(\hat{\Omega})}{\lambda_{\min}(\hat{\Omega})}} + 1\right) \|\sqrt{nh}(\tilde{\theta}_0 - \theta_0)\|.$$

Therefore, with

$$\tilde{c}_{1-\alpha,n}\left(\theta_0, \hat{\Sigma}, \hat{\Omega}\right) = c_{1-\alpha,n}\left(\theta_0, \hat{\Sigma}, \hat{\Omega}\right) + \frac{C_B}{\sqrt{\min_{k=1,\ldots,K_n} \hat{\Sigma}_{kk}}}\left(2 + \sqrt{\frac{\max_{k=1,\ldots,K_n} \hat{\Sigma}_{kk}}{\min_{k=1,\ldots,K_n} \hat{\Sigma}_{kk}}}\right)$$

we get

$$\lim_{n \to \infty} \inf_{P \in \mathcal{P}} P\left(T\left(\sqrt{nh_n}(\hat{\theta}_r - \theta_0), \hat{\Sigma}\right) \leq \tilde{c}_{1-\alpha,n}\left(\theta_0, \hat{\Sigma}, \hat{\Omega}\right)\right) \geq 1 - \alpha.$$

A.4 Computational details

We use two different starting values for each of our grid points. The first starting value is the restricted estimator. For the second starting value, we make use of the first steps of the algorithm recently proposed by Kaido et al. (2016). In particular, we follow several steps:

1. Calculate $c_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$ at $40d_\theta + 1$ randomly drawn points from Θ_R.

2. Approximate $c_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$ by a flexible auxiliary model which yields $c^A_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$ for any $\theta \in \Theta$ and $c^A_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$ is available in closed form.

3. Maximize/minimize $p_{K_n}(x_l)'\theta$ subject to $T(\sqrt{n}(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c^A_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$.

We draw the initial points from a normal distribution with a mean of $\hat{\theta}_r$ and a variance of $2 \cdot var(\hat{\theta}_{ur})$ and then project them onto the restricted parameter space. For each point, we calculate the critical value using $2,000$ simulation draws. The auxiliary model is a Gaussian process regression, just as in Kaido et al. (2016). We then obtain our starting values by solving optimization problems of the form $\max(\min)\, p_{K_n}(x_l)'\theta$ subject to $\theta \in \Theta_R$ and $T(\sqrt{n}(\hat{\theta}_r - \theta), \hat{\Sigma}) \leq c^A_{1-\alpha}(\theta, \hat{\Sigma}, \hat{\Omega})$.

Kaido et al. (2016) suggest using an iterative procedure, where an additional point is drawn in each iteration. Using this procedure instead of our direct approach yields essentially identical results in both the simulations and in the application, and their approach is much faster.

A.5 Spline results

In this section we present analogous results to those in Section 1.6 using quadratic splines as basis functions. Knots refers to the number of interior knots. Hence, 0 knots are equivalent to a quadratic function and with 2 knots we have 5 basis functions. If we have one interior knot, we choose it to be 0 and with two interior knots we take $-1/3$ and $1/3$.

Table 5: Coverage and width comparison for regression with splines

knots	c	cov_{ur}	cov_r	$width_{ur}$	$width_r$	% gains
	0	0.949	0.954	0.142	0.109	0.236
	2	0.947	0.963	0.142	0.121	0.143
0	4	0.948	0.960	0.142	0.129	0.091
	6	0.925	0.939	0.142	0.134	0.050
	8	0.910	0.910	0.142	0.137	0.028
	0	0.887	0.884	0.142	0.139	0.015
	0	0.951	0.975	0.172	0.128	0.258
	2	0.944	0.966	0.172	0.141	0.180
1	4	0.946	0.968	0.172	0.151	0.125
	6	0.950	0.956	0.172	0.157	0.086
	8	0.930	0.948	0.172	0.162	0.057
	10	0.942	0.947	0.172	0.165	0.037
	0	0.948	0.968	0.199	0.144	0.277
	2	0.945	0.973	0.199	0.157	0.211
2	4	0.948	0.966	0.199	0.168	0.157
	6	0.952	0.960	0.199	0.175	0.120
	8	0.935	0.945	0.199	0.180	0.094
	10	0.950	0.964	0.199	0.185	0.071

Table 6: Coverage and width comparison for NPIV with splines

knots	c	cov_{ur}	cov_r	$width_{ur}$	$width_r$	% gains
0	0	0.933	0.963	0.107	0.061	0.426
	5	0.931	0.949	0.107	0.079	0.257
	10	0.921	0.940	0.107	0.091	0.150
	20	0.821	0.815	0.107	0.101	0.049
	30	0.681	0.680	0.107	0.105	0.018
	40	0.426	0.426	0.107	0.106	0.002
	50	0.201	0.201	0.106	0.106	0.000
1	0	0.952	0.989	0.228	0.092	0.597
	5	0.951	0.977	0.229	0.113	0.506
	10	0.948	0.959	0.229	0.131	0.428
	20	0.947	0.945	0.228	0.157	0.315
	30	0.960	0.963	0.229	0.177	0.233
	40	0.952	0.953	0.228	0.191	0.160
	50	0.946	0.950	0.229	0.200	0.117
2	0	0.978	0.988	0.597	0.136	0.773
	5	0.972	0.988	0.595	0.171	0.715
	10	0.973	0.973	0.604	0.198	0.673
	20	0.966	0.961	0.611	0.233	0.621
	30	0.978	0.971	0.593	0.261	0.564
	40	0.975	0.965	0.603	0.287	0.525
	50	0.969	0.977	0.598	0.308	0.484

Appendix B

Simple Inference in First-Price Auctions

B.1 Tables and Figures

Functional	n	Root Mean Squared Error		
		Our Estimator	Method of Moments	Maximum Log-likelihood
Density at a Point	500	0.0159	0.0159	0.0160
	1,000	0.0111	0.0111	0.0111
	2,000	0.0079	0.0079	0.0079
Optimal Reserve Price	500	0.1158	0.1157	0.1150
	1,000	0.0820	0.0822	0.0813
	2,000	0.0574	0.0574	0.0572
Expected Revenue	500	0.0669	0.0626	0.0629
	1,000	0.0472	0.0440	0.0437
	2,000	0.0328	0.0313	0.0317

Table 7: This table presents the square-root of the mean squared error (rmse) for the indicated functionals for the three indicated estimators.

Empirical Coverage : Our Proposed Estimator

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

Table 8: 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.

Root Mean Squared Error

Functional	n	Our Estimator			M+S (2012)	GPV (2000)
		$K_n = 1$	$K_n = 2$	$K_n = 3$		
Density at a Point	500	0.0084	0.0117	0.0159	0.0364	0.0364
	1,000	0.0065	0.0066	0.0111	0.0297	0.0311
	1,500	0.0057	0.0066	0.0093	0.0250	0.0272
	2,000	0.0053	0.0055	0.0079	0.0230	0.0251
Optimal Reserve Price	500	0.0942	0.1166	0.1158	0.2178	0.2761
	1,000	0.0674	0.0638	0.0820	0.1681	0.2282
	1,500	0.0549	0.0638	0.0661	0.1499	0.1980
	2,000	0.0510	0.0549	0.0574	0.1341	0.1752
Expected Revenue	500	0.0641	0.0668	0.0669	0.1623	0.3101
	1,000	0.0439	0.0384	0.0472	0.1317	0.1615
	1,500	0.0355	0.0386	0.0360	0.1205	0.1049
	2,000	0.0321	0.0322	0.0328	0.1102	0.0721

Table 9: 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).

Functional	n	Our Estimator		M+S (2012)		Ma et al (2016)	
		Coverage	Avg CS Size	Coverage	Avg CS Size	Coverage	Avg CS Size
Density at a Point	500	95.3%	0.064	89.8%	0.127	92.4%	0.136
	1,000	95.2%	0.045	92.6%	0.111	92.6%	0.117
	1,500	94.6%	0.036	94.3%	0.104	93.8%	0.108
	2,000	94.6%	0.031	95.0%	0.097	95.2%	0.102
	4,000	95.3%	0.022	97.3%	0.086	94.3%	0.087
Optimal Reserve Price	500	94.2%	0.437	79.8%	0.654	N/A	N/A
	1,000	94.1%	0.311	86.6%	0.525	N/A	N/A
	1,500	94.1%	0.257	87.1%	0.471	N/A	N/A
	2,000	93.2%	0.223	89.9%	0.445	N/A	N/A
	4,000	95.1%	0.157	93.6%	0.374	N/A	N/A
Expected Revenue	500	94.8%	0.253	N/A	N/A	N/A	N/A
	1,000	93.8%	0.178	N/A	N/A	N/A	N/A
	1,500	96.2%	0.145	N/A	N/A	N/A	N/A
	2,000	94.9%	0.126	N/A	N/A	N/A	N/A
	4,000	93.8%	0.089	N/A	N/A	N/A	N/A

Table 10: 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), 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.

p	2	3	4	5	6	7	8	9	Total
Count	324	246	189	127	105	75	41	22	1,129

Table 11: 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 π described in section 2.6

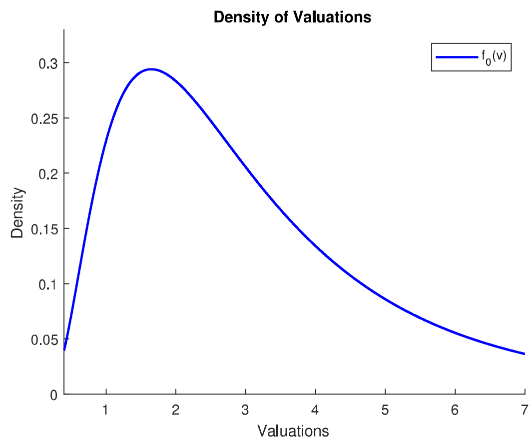


Figure 7: This figure plots the parametric distribution of valuations discussed in section 2.2.2

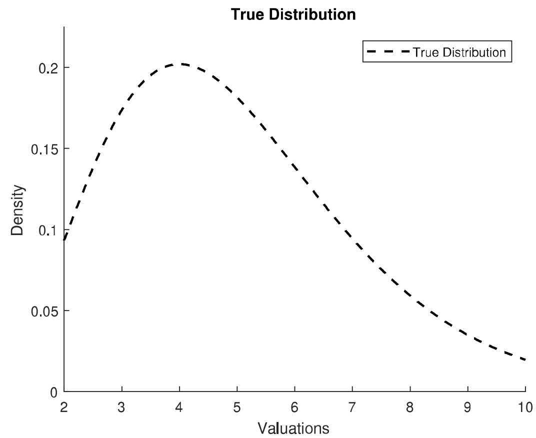


Figure 8: This figure contains the true density of the Monte Carlo simulation presented in section 2.5

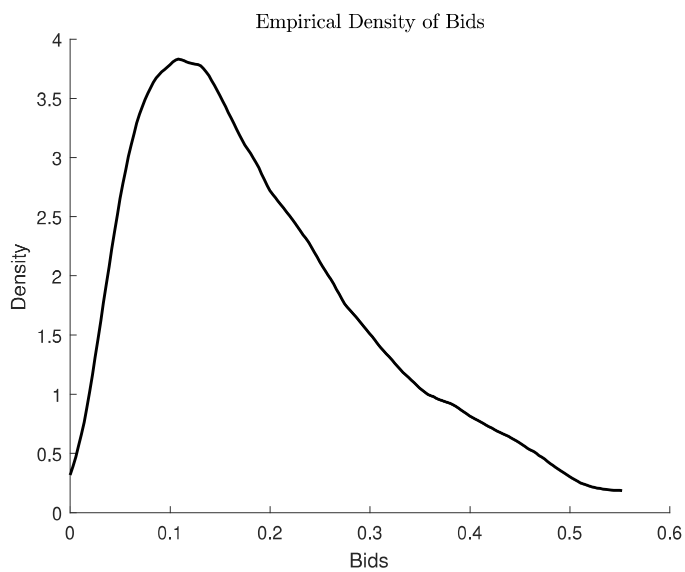


Figure 9: This figure illustrates a kernel-smoothed estimate of the density of bids for the application to USFS timber auctions.

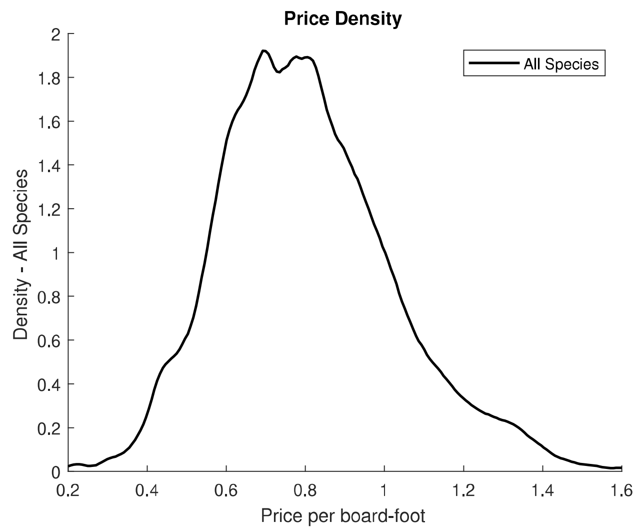


Figure 10: This figure illustrates a kernel-smoothed estimate of the density of all appraised prices for all species of timber.

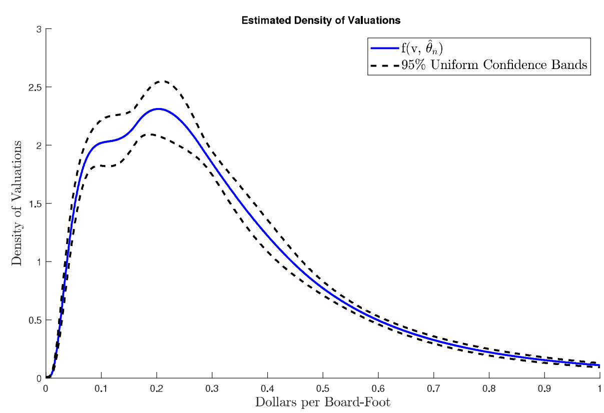


Figure 11: This figure contains our point-estimate for the density of valuations using quadratic splines with four knots. The dashed lines correspond to 95% uniform confidence bands for the true density of valuations.

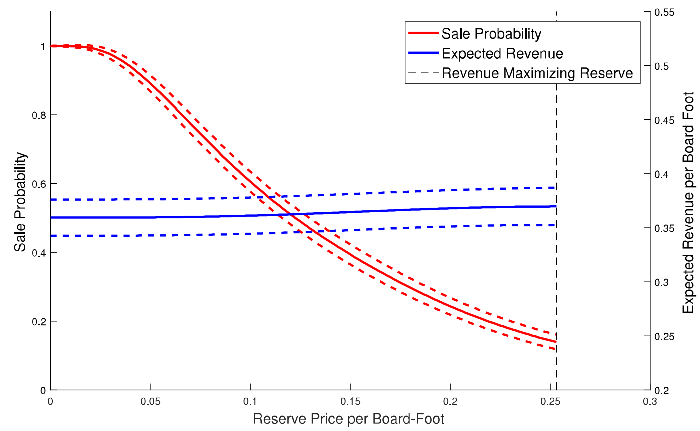


Figure 12: This figure illustrates the trade-off between higher revenues and a lower probability of selling timber as the reserve price is increased from no reserve price to the revenue-maximizing reserve price. The decreasing function (and the associated dashed lines) correspond to the point estimate 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.

B.2 Mathematical Proofs

B.2.1 Proof of Theorem 2.1

Let Z_n denote the normal random variable described in Assumption (2.2). 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 the expression for our proposed estimator

$$\begin{aligned}
\sqrt{n}\,(\hat{\theta}_n - \theta_0) &= \sqrt{n}\left(\hat{\theta}_{\text{mm}} - \theta_0\right) - \left(\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{n}\,\frac{\partial Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta} \\
&= \sqrt{n}\,(\theta_n - \theta_0) + \left(\left(\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1} - \left(\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta\partial\theta'}\right)^{-1}\right)\sqrt{n}\left(\frac{\partial Q_n(\theta_0)}{\partial\theta}\right) \\
&\quad + \left(I_{d_\theta} - \left(\frac{\partial^2 Q_n(\hat{\theta}_{\text{mm}})}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2 Q_n(\theta^+)}{\partial\theta\partial\theta'}\right)\sqrt{n}\,(\hat{\theta}_{\text{mm}} - \theta_0)
\end{aligned}$$

where the second line follows from a mean-value expansion where θ^+ is the intermediate value with $\|\theta^+ - \theta_0\| \leq \|\hat{\theta}_{\text{mm}} - \theta_0\|$. The third line uses the definition $\sqrt{n}\,(\theta_n - \theta_0) = \left(\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta}$. Therefore, by the triangle inequality and properties of the spectral norm, we can re-arrange the above equation to get

$$\begin{aligned}
&\left\|\sqrt{n}\,(\hat{\theta}_n - \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)\right\| \\
&\quad\leq \left\|\left(\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1} - \left(\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}\right)^{-1}\right\|_S\left\|\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta}\right\| \\
&\quad\quad + \left\|I_{d_\theta} - \left(\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2 Q_n(\theta^+)}{\partial\theta\partial\theta'}\right\|_S\left\|\sqrt{n}\,(\hat{\theta}_r - \theta_0)\right\| \\
&\quad\leq \left\|\left(\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\right\|_S\left\|\left(\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}\right)^{-1}\right\|_S\left\|\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}\right\|_S\left\|\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta}\right\| \\
&\quad\quad + \left\|\left(\frac{\partial^2 Q_n(\hat{\theta}_{\text{r}})}{\partial\theta\partial\theta'}\right)^{-1}\right\|\left\|\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}\right\|\left\|\sqrt{n}\,(\hat{\theta}_r - \theta_0)\right\| \qquad \text{(B.1)}
\end{aligned}$$

where the second inequality uses the fact that for any invertible matrices A and B

$$\|A^{-1} - B^{-1}\|_S \|A^{-1}(B - A)B^{-1}\|S_{\leq} \|A^{-1}\|_S \|B^{-1}\|_S \|A - B\|_S$$

and

$$\|I - A^{-1}B\|_S = \|A^{-1}(A-B)\|_S \leq \|A^{-1}\|_S\|A-B\|_S.$$

Notice, by Assumptions (2.2) and (2.5) combined with lemma (B.2) and the fact that by Markov's inequality

$$\begin{aligned}
\left\|\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta}\right\| &\leq \lambda_{\max}\left(\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right)\left\|\left(\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right)^{-1}\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial\theta}\right\| \\
&\leq O_p(1)\big\|Z_n\big\| + o_p(1) \\
&= O_p(\sqrt{K_n})
\end{aligned}$$

where the second inequality uses Assumptions (2.2) and (2.5). Furthermore, by Assumptions (2.1) and (2.5), we have

$$\begin{aligned}
\big\|\sqrt{n}\,(\hat{\theta}_r - \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)\big\| &\leq \lambda_{\min}(J_n)^{-\frac{1}{2}}\big\|\sqrt{n}\,(\hat{\theta}_r - \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)\big\|_{J_n} \\
&\leq \lambda_{\min}(J_n)^{-\frac{1}{2}}\big\|\sqrt{n}\,(\theta_n - \theta_0)\big\|_{J_n} + \lambda_{\min}(J_n)^{-\frac{1}{2}}\lambda_{\max}(J_n)^{\frac{1}{2}}\|r_n\| \\
&\leq \lambda_{\min}(J_n)^{-\frac{1}{2}}\big\|Z_n\big\|_{J_n} + o_p(\varepsilon_n) \\
&\leq \lambda_{\max}(J_n)^{\frac{1}{2}}\lambda_{\min}(J_n)^{-\frac{1}{2}}\lambda_{\max}(\Sigma_n)^{\frac{1}{2}}\|\Sigma_n^{-\frac{1}{2}}Z_n\| + o_p(\varepsilon_n) \\
&= O_p\big(\sqrt{K_n}\big)
\end{aligned}$$

where the second inequality uses Assumption (2.1) and the last line uses Assumption (2.5), Assumption (2.2) and Markov's inequality. Using the triangle inequality gives

$$\big\|\sqrt{n}\,(\hat{\theta}_r - \theta_0)\big\| \leq \big\|\sqrt{n}\,(\hat{\theta}_r - \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)\big\| + \big\|\sqrt{n}\,(\theta_n - \theta_0)\big\| = O_p\big(\sqrt{K_n}\big) \tag{B.2}$$

uniformly over $\mathcal{P}$. Plugging these bounds into equation (B.1) and using Assumption (2.4) gives

$$\left\|\sqrt{n}\,(\hat{\theta}_n - \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)\right\| = o_p(\varepsilon_n). \tag{B.3}$$

Therefore, by the triangle inequality and Assumption (2.2), we get

$$\|\sqrt{n}\,(\hat{\theta}_n - \theta_0) - Z_n\| \leq \|\sqrt{n}\,(\hat{\theta}_n - \theta_0) - \sqrt{n}\,(\theta_n - \theta_0)\| + \|\sqrt{n}\,(\theta_n - \theta_0) - Z_n\| = o_p(\varepsilon_n).$$

In the remainder of the proof, we show that this approximation error and the errors 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 Assumption (2.2), Assumption (2.6) and the triangle-inequality we have

$$\begin{aligned}
&\Big|T(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n)-T(Z_n,\Sigma_n)\Big| \\
&\qquad\leq \Big|T(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n)-T(Z_n,\hat{\Sigma}_n)\Big| + \Big|T(Z_n,\hat{\Sigma}_n)-T(Z_n,\Sigma_n)\Big| \\
&\qquad\leq C\big\|\sqrt{n}\,(\hat{\theta}_n-\theta_0)-Z_n\big\| + C\big\|Z_n\big\|\big\|\hat{\Sigma}_n-\Sigma_n\big\|_S \\
&\qquad= o_p(\varepsilon_n) \qquad\qquad\qquad\qquad \text{(B.4)}
\end{aligned}$$

where the last line follows from Markov's inequality, Assumption (2.3) and the observation that

$$\big\|Z_n\big\| = \big\|\Sigma_n^{\frac{1}{2}}\Sigma_n^{-\frac{1}{2}}Z_n\big\| \leq \lambda_{\max}\big(\Sigma_n\big)^{\frac{1}{2}}\big\|N(0,I_{K_n})\big\| = \lambda_{\max}\big(\Sigma_n\big)^{\frac{1}{2}}O_p(\sqrt{K_n}).$$

Letting $\mathcal{A}_n$ denote the event that $\Big|T(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n)-T(Z_n,\Sigma_n)\Big| < \frac{1}{2}\varepsilon_n$, the above shows $P(\mathcal{A}_n^C)=o(1)$. Therefore, by basic probability rules

$$\begin{aligned}
&P\Big(T(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)\Big) \\
&\qquad\leq P\Big(T(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)\,,\,\mathcal{A}_n\Big) + P\Big(\mathcal{A}_n^C\Big) \\
&\qquad\leq P\Big(T(Z_n,\Sigma_n)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)+\frac{1}{2}\varepsilon_n\Big)+o(1). \qquad\qquad \text{(B.5)}
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&P\Big(T\big(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n\big)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)\Big)\\
&\qquad\geq P\Big(T\big(\sqrt{n}\,(\hat{\theta}_n-\theta_0),\hat{\Sigma}_n\big)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)\,,\,\mathcal{A}_n\Big)\\
&\qquad\geq P\Big(T\big(Z_n,\Sigma_n\big)+\frac{1}{2}\varepsilon_n\leq c_{1-\alpha,n}(\hat{\Sigma}_n)\,,\,\mathcal{A}_n\Big)\\
&\qquad\geq P\Big(T\big(Z_n,\Sigma_n\big)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)-\frac{1}{2}\varepsilon_n\Big)-P(\mathcal{A}_n^C)\\
&\qquad= P\Big(T\big(Z_n,\Sigma_n\big)\leq c_{1-\alpha,n}(\hat{\Sigma}_n)-\frac{1}{2}\varepsilon_n\Big)-o(1) \qquad\qquad (B.6)
\end{aligned}
$$

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.[1]. 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 $\varepsilon_n=0$ and $\hat{\Sigma}_n=\Sigma_n$, the previous displays would provide the result. In the next part of the proof, we establish that the approximation error in using $\hat{\Sigma}_n$ to compute the critical value is asymptotically small. Specifically, we show that for any arbitrarily small $\delta_q>0$ we have,

$$c_{1-\alpha,n}(\hat{\Sigma}_n)\geq c_{1-\alpha-\delta_q,n}(\Sigma_n)-\frac{1}{2}\varepsilon_n \qquad (B.7)$$

and

$$c_{1-\alpha,n}(\hat{\Sigma}_n)\leq c_{1-\alpha+\delta_q,n}(\Sigma_n)+\frac{1}{2}\varepsilon_n. \qquad (B.8)$$

To establish the first result, first notice by Assumption (2.3), Assumption (2.6) and Markov's inequality,

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

[1]Note, the inequality uses the fact that for any two measurable events A,B, we have $P(A)=P(A\cap B)+P(A\cap B^c)$ which implies $P(A\cap B)=P(A)-P(A\cap B^c)\geq P(A)-P(B^c)$

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

$$\begin{aligned} 1-\alpha &= P\Big(T(Z_n, \hat{\Sigma}_n) \le c_{1-\alpha,n}(\hat{\Sigma})\Big) \\ &\le P\Big(T(Z_n, \Sigma_n) \le c_{1-\alpha,n}(\hat{\Sigma}) + \frac{1}{2}\varepsilon_n \,,\, \mathcal{B}_n\Big) + P\Big(\mathcal{B}_n^C\Big) \\ &\le P\Big(T(Z_n, \Sigma_n) \le c_{1-\alpha,n}(\hat{\Sigma}) + \frac{1}{2}\varepsilon_n\Big) + o(1) \end{aligned}$$

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 δ_q uniformly over $P \in \mathcal{P}$ so that the previous display shows

$$1-\alpha-\delta_q \le P\Big(T(Z_n, \Sigma_n) \le c_{1-\alpha,n}(\hat{\Sigma}) + \frac{1}{2}\varepsilon_n\Big)$$

so that $c_{1-\alpha-\delta_q,n}(\Sigma) \le c_{1-\alpha,n}(\hat{\Sigma}) + \frac{1}{2}\varepsilon_n$, which establishes the first result. To establish equation (B.8), we follow a similar argument to get

$$\begin{aligned} 1-\alpha &= P\Big(T(Z_n, \Sigma_n) \le c_{1-\alpha,n}(\Sigma_n)\Big) \\ &\ge P\Big(T(Z_n, \hat{\Sigma}_n) + \frac{1}{2}\varepsilon_n \le c_{1-\alpha,n}(\Sigma_n) \,,\, \mathcal{B}_n\Big) \\ &\ge P\Big(T(Z_n, \hat{\Sigma}_n) \le c_{1-\alpha,n}(\Sigma_n) - \frac{1}{2}\varepsilon_n\Big) - P\Big(\mathcal{B}_n^C\Big) \\ &= P\Big(T(Z_n, \hat{\Sigma}_n) \le c_{1-\alpha,n}(\Sigma_n) - \frac{1}{2}\varepsilon_n\Big) - o(1) \end{aligned}$$

where we can make the $o(1)$ term uniformly (over $\mathcal{P}$) smaller than δ_q. Hence, we have

$$1-\alpha+\delta_q \ge P\Big(T(Z_n, \hat{\Sigma}_n) \le c_{1-\alpha,n}(\Sigma_n) - \frac{1}{2}\varepsilon_n\Big)$$

so that $c_{1-\alpha+\delta_q}(\hat{\Sigma}) \ge c_{1-\alpha,n}(\Sigma_n) - \frac{1}{2}\varepsilon_n$ which establishes equation (B.8). Combining equations (B.6) and (B.7) gives

$$\begin{aligned} P\Big(T(\sqrt{n}\,(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \le c_{1-\alpha,n}(\hat{\Sigma}_n) + \varepsilon_n\Big) &\ge P\Big(T(Z_n, \Sigma_n) \le c_{1-\alpha,n}(\hat{\Sigma}_n) + \frac{1}{2}\varepsilon_n\Big) - o(1) \\ &\ge P\Big(T(Z_n, \Sigma_n) \le c_{1-\alpha-\delta_q,n}(\Sigma_n)\Big) - o(1) \\ &= 1-\alpha-\delta_q - o(1). \end{aligned}$$

As we can make the $o(1)$ term arbitrarily small uniformly over $\mathcal{P}$ and the choice of δ_q is arbitrary, we have

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} P\Big(T(\sqrt{n}\,(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1-\alpha,n}(\hat{\Sigma}_n) + \varepsilon_n\Big) \geq 1-\alpha$$

which is the first conclusion of the theorem. Additionally, for any δ_q equations (B.6) and equation (B.7) imply

$$\begin{aligned} P\Big(T(\sqrt{n}(\hat{\theta}_n - \theta_0), \hat{\Sigma}_n) \leq c_{1-\alpha,n}(\hat{\Sigma}_n)\Big) &\geq P\Big(T(Z_n, \Sigma_n) \leq c_{1-\alpha,n}(\hat{\Sigma}) - \frac{1}{2}\varepsilon_n\Big) - o(1) \\ &\geq P\Big(T(Z_n, \Sigma_n) \leq c_{1-\alpha-\delta_q,n}(\Sigma) - \varepsilon_n\Big) - o(1). \end{aligned}$$

Similarly, by equations (B.5) and (B.8) we have

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

Therefore, if Assumption (2.7) holds, then for each δ_q, we have

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

which establishes the theorem.

B.2.2 Supplementary Lemmas

Lemma B.1. ***Useful Properties of Eigenvalues***

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

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

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

Proof. By the definition of the smallest eigenvalues,

$$\begin{aligned}\left|\lambda_{\min}(A_n) - \lambda_{\min}(\hat{A}_n)\right| &= \left|\inf_{\|x\|=1} x'A_n x - \inf_{\|x\|=1} x'\hat{A}_n x\right| \\ &\leq \sup_{\|x\|=1} |x'(A_n - \hat{A}_n)x| \\ &\leq \|A_n - \hat{A}_n\|_S \\ &\leq \|A_n - \hat{A}_n\|.\end{aligned}$$

Similarly, by the definition of the largest eigenvalues,

$$\begin{aligned}\left|\lambda_{\max}(A_n) - \lambda_{\max}(\hat{A}_n)\right| &= \left|\sup_{\|x\|=1} x'A_n x - \sup_{\|x\|=1} x'\hat{A}_n x\right| \\ &\leq \sup_{\|x\|=1} |x'(A_n - \hat{A}_n)x| \\ &\leq \|A_n - \hat{A}_n\|_S \\ &\leq \|A_n - \hat{A}_n\|.\end{aligned}$$

□

Lemma B.2. ***Plug-in Second Derivative is Positive Definite***

Under Assumptions (2.1), (2.2), (2.4) and (2.5), the eigenvalues of $\frac{\partial^2 Q_n(\hat{\theta}_{mm})}{\partial\theta\partial\theta'}$ *are bounded and bounded away from zero*

Proof. To begin, notice

$$\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'} = \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'} + \left(\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right). \tag{B.9}$$

From Assumption (2.4) we have

$$\left\|\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right\| = O_p\left(\sqrt{n}K_n\right) = o_p(\varepsilon_n).$$

As $\left\|\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}\right\| = o_p(1)$ and the Frobeninus norm is the square-root of the sum of square eigenvalues, the previous display shows the absolute value of the eigenvalues of $\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'} -$

$\frac{\partial^2 Q_n(\theta_0)}{\partial\theta\partial\theta'}$ 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 (B.9) shows the smallest eigenvalue of $\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}$ 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 $\frac{\partial^2 Q_n(\hat{\theta}_r)}{\partial\theta\partial\theta'}$ is also strictly larger than zero with probability approaching one. □

Lemma B.3. ***Plug-in Estimate of*** Σ_n ***is Positive Definite***

Under Assumptions (2.3) and (2.5), $\hat{\Sigma}_n$ *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 Σ_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. □

B.2.3 Proof of Theorem 2.2

Before providing the proof of the theorem, we formally re-state the assumptions and the statement of the theorem. To state the assumptions let $\zeta_L \equiv \sup_{v\in[\underline{v},\overline{v}]} \|P(v)\|$ and $\mathcal{P}$ be the class of distributions satisfying the following assumptions. There exists constants $0 < C\,, \overline{p} < \infty$ not depending on L or P such that

Assumption B.1. ***Model Specification*** *The data* $\{\boldsymbol{b}_l\}_{l=1}^{L}$ *satisfies the following properties*

i) Independent Private Values There exists an independent and identically distributed sequence $\{v_j\}_{j=1}^{n}$ *(*$n \equiv \sum_{l=1}^{L} p_l$*) with* v_j *supported on* $[\underline{v}, \overline{v}]$ *with density* $f_0(v)$ *such that for* *each* $l \in \{1, ..., L\}$*,* $\boldsymbol{b}_l = (b_{1l}, ..., b_{p_l l})$ *with*

$$b_{il} = v_{\iota(i,l)} - \frac{\int_{\underline{v}}^{v_{\iota(i,l)}} F_0(t)^{p_l - 1} dt}{F_0(v_{\iota(i,l)})^{p_l - 1}}$$

where $\iota(1,1) = 1 \, , \, \ldots \, , \, \iota(p_l, 1) = p_l \, , \, \iota(1,2) = p_l + 1 \, , \, \ldots \, , \, \iota(p_L, L) = n$*.*

ii) Distribution of Number of Participants: $\{p_l\}_{l=1}^{L}$ *is a sequence of independent and identically distributed random variables with support contained in* $\{2, 3, .., \overline{p}\}$ *for some* $\overline{p} < \infty$*.*

iii) Exogenous Entry: For all $l \in \{1, ...L\}$*,* $v_{il} \perp p_l$ *for all* $i \in \{1, .., p_l\}$*.*

Assumption B.2. ***Sieve Approximation Error*** *There exists a sequence* $\theta_{0,L} \in \Theta_{K_L}$ *such that for some* $\gamma > 0$

$$\sup_{v \in [\underline{v}, \overline{v}]} |f(v, \theta_0) - f_0(v)| \leq C K_L^{-\gamma} \quad \text{and} \quad \overline{b}(\theta_0, p) \geq \overline{b}_0(p)$$

where $\overline{b}_0(p) = \overline{v} - \int_{\underline{v}}^{\overline{v}} F_0(t)^{p-1} dt$ *denotes the largest bid rationalized by* $f_0(v)$*.*

Assumption B.3. ***Density Bounds*** *For all* $\theta \in \Theta$ *and* $v \in [\underline{v}, \overline{v}]$

$$\frac{1}{C} < f_0(v) \, , \, f(v, \theta) \, , \, f'(v, \theta) \, , \, \phi(v) < C.$$

Assumption B.4. ***Rate Conditions*** *The following rate conditions hold*

$$\frac{K_L^3 \zeta_L^4}{L} = o_p(\varepsilon_L^6) \, , \, \frac{K_L^4 \zeta_L^2}{L} = o_p(\varepsilon_L^6) \, , \, L K_L^{1-2\gamma} = o_p(\varepsilon_L).$$

Assumption B.5. ***Bounded Eigenvalues*** *The eigenvalues of the matrices*

$$\mathbb{E}_0 \left(m(\boldsymbol{b}, \theta_0)' m(\boldsymbol{b}, \theta_0) \right) \quad \text{and} \quad \mathbb{E}_0 \left(\frac{dm(\boldsymbol{b}, \theta_0)}{d\theta'} \right) \mathbb{E}_0 \left(\frac{dm(\boldsymbol{b}, \theta_0)}{d\theta'} \right)'$$

lie in the interval $\left[\frac{1}{C} \, , \, C \right]$

Theorem B.2. *Assumptions (B.1) - (B.5) imply Assumptions (2.2) - (2.5) in section 2.3.3.*

Proof. For organizational purpose, the proof of this theorem is broken up into several compartmentalized lemmas appearing below in this section. Lemma (B.4) establishes that Assumptions (B.1) - (B.5) imply Assumption (2.2). Lemma (B.5) establishes that Assumptions (B.1) - (B.5) are sufficient for Assumption (2.3). Lemma (B.6) establishes that Assumptions (B.1) - (B.5) imply Assumption (2.4). Finally, lemma (B.7) establishes the final statement that Assumptions (B.3) - (B.5) implies Assumption (2.5). □

Lemma B.4. ***Asymptotic Normality of Infeasible Estimator*** *Let* $\theta_L = \theta_{0,L} + \frac{1}{\sqrt{L}} J_L^{-1} D_L$ *where*

$$D_L = -\sqrt{L}\,\frac{dQ_L(\theta_{0,L})}{\partial\theta} \text{ and } J_L = \frac{d^2 Q_L(\theta_{0,L})}{d\theta d\theta'}$$

with $Q_L(\theta)$ *and* $\theta_{0,L}$ *defined in Assumption (B.2). Under Assumptions (B.1) - (B.4), there exists a sequence of random variables* $\{\Sigma_L, Z_L\}_{L=1}^{\infty}$ *with* $Z_L \sim N(0, \Sigma_L)$ *such that* $\|\sqrt{L}\,(\theta_L - \theta_{0,L}) - Z_L\| = o_p(\varepsilon_n)$.

Proof. Re-arranging the definition of θ_L gives $\sqrt{L}\,(\theta_L - \theta_{0,L}) = J_L^{-1} D_n$. Let,

$$Z_L = -\Big(\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta'}\Big)\Big)^{-1}\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big) S_L$$

where S_L is the random variable defined in lemma (B.8). Note,

$$D_L = \Big(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_{0,L})}{d\theta}\Big)\Big(\frac{-1}{\sqrt{L}}\sum_{l=1}^{L} m(\boldsymbol{b}_l, \theta_{0,L})\Big).$$

Therefore,

$$\begin{aligned}
&\Big\|D_L - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big)\Big(-S_L\Big)\Big\| \\
&\qquad \le \Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_{0,L})}{d\theta}\Big(S_L - \frac{1}{\sqrt{L}}\sum_{l=1}^{L} m(\boldsymbol{b}_l, \theta_{0,L})\Big)\Big\| \qquad \text{(B.10)} \\
&\qquad\quad + \Big\|\Big(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_{0,L})}{d\theta} - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big)\Big) S_L\Big\|. \qquad \text{(B.11)}
\end{aligned}$$

To bound the first value, notice

$$\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}\Big(S_L-\frac{1}{\sqrt{L}}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_{0,L})\Big)\Big\| \\ \leq \Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}\Big\|_S\Big\|S_l-\frac{1}{\sqrt{L}}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_{0,L})\Big\|.$$

By the reverse triangle inequality, and lemma (B.9) and Assumption (B.4),

$$\Bigg|\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}\Big\|_S-\Big\|\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|_S\Bigg|=o_p(1).$$

As $\|A\|_S=\lambda_{\max}(AA')$, $\Big\|\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|_S=O(1)$ by Assumption (B.5). Therefore, $\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}\Big\|_S=O_p(1)$. Combining the last several inequalities with lemma (B.8) gives

$$\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}\Big(S_L-\frac{1}{\sqrt{L}}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_{0,L})\Big)\Big\|=O_p(1)o_p(\varepsilon_L)=o_p(\varepsilon_L). \tag{B.12}$$

To bound the second term in the sum of equation (B.10), note by Markov's inequality and Assumption (B.5)

$$\Big\|S_L\Big\|\leq\sqrt{\lambda_{\max}(V_L)}\Big\|N(0,I_{K_L})\Big\|=O_p(\sqrt{K_L}).$$

Therefore, by lemma (B.9) and the previous derivation,

$$\Big\|\Big(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}-\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big)S_L\Big\| \\ \leq\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}-\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|\Big\|S_L\Big\| \\ =O_p\Big(\frac{\zeta_L^2}{\sqrt{L}}\Big)O_p(\sqrt{K_L})=o_p(\varepsilon_L) \tag{B.13}$$

where the last equality follows from Assumption (B.4). Using expressions (B.12) and (B.13) with equation (B.10) yields

$$\Big\|D_L-\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big(-S_L\Big)\Big\|=o_p(\varepsilon_L) \tag{B.14}$$

To establish the result, notice by the definition of Z_L and the triangle inequality

$$\begin{aligned}\Big\|J_L^{-1}D_L - Z_L\Big\| &\leq \Big\|J_L^{-1}\Big(D_L - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big(-S_L\Big)\Big\| \\ &+ \Big\|\Big(\Big(\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta'}\Big)\Big)^{-1} - J_L^{-1}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)S_L\Big\|, \qquad \text{(B.15)}\end{aligned}$$

so it suffices to provide a bound for each of the two terms. Starting with the first, notice

$$\begin{aligned}&\Big\|J_L^{-1}\Big(D_L - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big(-S_L\Big)\Big)\Big\| \\ &\leq \Big\|J_L^{-1}\Big\|_S\Big\|D_L - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big(-S_L\Big)\Big\| \\ &= o_p(\varepsilon_L) \qquad \text{(B.16)}\end{aligned}$$

where the equality follows from equation (B.14) and the fact that J_L has bounded eigenvalues with probability approaching one by the proof of lemma (B.5). For the second term, notice

$$\begin{aligned}&\Big\|\Big(\Big(\mathbb{E}_0\Big(\frac{dm\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\mathbb{E}_0\Big(\frac{dm\boldsymbol{b},\theta_{0,L})}{d\theta'}\Big)\Big)^{-1} - J_L^{-1}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)S_L\Big\| \\ &\leq \Big\|J_L^{-1}\Big\|\Big\|\Big(\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta'}\Big)\Big)^{-1}\Big\|_S \\ &\quad \times \Big\|J_L - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta'}\Big)\Big\|_S\Big\|\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|_S\Big\|S_L\Big\| \\ &= o_p(\varepsilon_L) \qquad \text{(B.17)}\end{aligned}$$

where the first inequality follows from the definition (and sub- multiplicative property) of the spectral norm and the observation that for invertible matrices A and B,

$$\|A^{-1} - B^{-1}\| = \|A^{-1}(B - A)B^{-1}\| \leq \|A^{-1}\|_S\|B^{-1}\|_S\|A - B\|.$$

The last equality in equation (B.17) follows from equation (B.21) in the proof of lemma (B.5) and the previous observation that $\|S_L\| = o_p(\varepsilon_L K_L^{\frac{1}{2}})$. Combining the last several equations yield

$$\Big\|J_L^{-1}D_L - Z_L\Big\| = o_p(\varepsilon_L)$$

where $Z_L \sim N(0, \Sigma_L)$ with

$$\Sigma_L \equiv \Big(\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big)\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big)\Big)^{-1}\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b}, \theta_{0,L})}{d\theta}\Big)V(m(\boldsymbol{b}, \theta_{0,L}))^{\frac{1}{2}}.$$

□

Lemma B.5. ***Covariance Estimator*** *Let* $\hat{\Sigma}_L = \hat{J}_L^{-1}\hat{V}_D\hat{J}_L^{-1}$ *where* $\hat{J}_L = \frac{d^2Q_L(\hat{\theta}_{mm})}{d\theta d\theta'}$ *and*

$$\hat{V}_D = \left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \hat{\theta}_{mm})}{d\theta}\right)\left(\frac{1}{L}\sum_{l=1}^{L}m(\boldsymbol{b}_l, \hat{\theta}_{mm})m(\boldsymbol{b}_l, \hat{\theta}_{mm})'\right)\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \hat{\theta}_{mm})}{d\theta}\right).$$

Further, let $\Sigma_L = J^{-1}V_DJ$ *where* $V_D = Var(D_L)$ *with* D_L *defined in equation (2.9) and*

$$J = \mathbb{E}_0\left(\frac{dm(\boldsymbol{b}, \theta_0)}{d\theta'}\right)\mathbb{E}_0\left(\frac{dm(\boldsymbol{b}, \theta_0)}{d\theta'}\right)'.$$

If Assumptions (2.8) - (2.5) hold,

$$\left\|\Sigma_L - \hat{\Sigma}_L\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right).$$

Proof. Note, by the triangle inequality

$$\begin{aligned}\left\|\Sigma_L - \hat{\Sigma}_L\right\| &= \left\|J^{-1}V_DJ^{-1} - \hat{J}_L^{-1}\hat{V}_D\hat{J}_L^{-1}\right\| \\ &\leq \left\|J^{-1}\right\|_S\left\|V_D\right\|_S\left\|J^{-1} - \hat{J}^{-1}\right\|_S + \left\|J^{-1}\right\|_S\left\|\hat{J}_L^{-1}\right\|_S\left\|V_D - \hat{V}_D\right\| \\ &\quad + \left\|J^{-1}\right\|_S\left\|\hat{V}_D\right\|_S\left\|J^{-1} - \hat{J}^{-1}\right\|_S \end{aligned} \tag{B.18}$$

therefore, it suffices to show each of these terms are $o_p(\varepsilon_L K^{-\frac{1}{2}})$. First, we establish $\|\hat{J}_L^{-1}\|_S$, $\|J^{-1}\|_S$, $\|V_D\|_S$ and $\|\hat{V}_D\|_S$ are all bounded and bounded away form zero. Assumption (B.5) immediately implies $\|J^{-1}\|_S$ is bounded and bounded away from zero. By Lemma (B.6),

$$\left\|\hat{J}_L - \frac{d^2Q_L(\theta_0)}{d\theta d\theta'}\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right) \tag{B.19}$$

as $\|\hat{\theta}_{\text{mm}} - \theta_0\| = \frac{1}{\sqrt{L}}\|\sqrt{L}\,(\hat{\theta}_{\text{mm}} - \theta_0)\| = o_p\left(\varepsilon_L K^{-\frac{1}{2}}\right)$. Also by lemma (B.6),

$$\left\|\frac{d^2Q_L(\theta_0)}{d\theta d\theta'} - \left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_0)}{d\theta}\right)\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_0)}{d\theta}\right)'\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right). \tag{B.20}$$

By lemma (B.9),

$$\left\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}-\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\right\|=o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right).$$

Therefore, by Weyl's the eigenvalues of $\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}$ are within a shrinking neighborhood of the eigenvalues of $\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta}\right)$, which (in absolute value) are bounded and bounded from zero. Hence, by properties of eigenvalues, $\left\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right\|_S$ and $\left\|\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\right\|_S$ are bounded and bounded from zero. Therefore, by lemma (B.9),

$$\begin{aligned}
&\left\|\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta}\right)\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta}\right)'-\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)'\right\| \\
&\qquad\leq\left(\left\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta'}\right\|_S+\left\|\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\right\|_S\right) \\
&\qquad\quad\times\left\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta'}-\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\right\| \\
&\qquad=o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)
\end{aligned}\tag{B.21}$$

Hence, by the triangle inequality and equations (B.19) - (B.21),

$$\left\|\hat{J}_L-\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)'\right\|=o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right).\tag{B.22}$$

By using this expression along with Weyl's inequality and Assumption (B.5), the eigenvalues of $\hat{J}_L$ are bounded and bounded away form zero with probability approaching one. Moreover, $V_D=J^{-1}V_mJ^{-1}$ where $V_m=Var(m(\boldsymbol{b},\theta_0))$, so that by Assumption (B.5), the eigenvalues of V_D are bounded and bounded away from zero. Therefore, notice,

$$\begin{aligned}
\left\|\hat{J}_L^{-1}\hat{V}_m\hat{J}_L^{-1}-V_D\right\|_S &\leq \left\|\hat{J}_L^{-1}\right\|_S^2\left\|\hat{V}_m-V_m\right\|+\left\|\hat{J}_L^{-1}\right\|_S\left\|V_m\right\|_S\left\|\hat{J}_L^{-1}-J_L^{-1}\right\|_S \\
&\quad+\left\|J_L\right\|\left\|V_m\right\|\left\|\hat{J}_L^{-1}-J^{-1}\right\|
\end{aligned}$$

By lemma (B.25) and the fact that $\|\hat{J}_L^{-1}\|_S$ and $\|J^{-1}\|_S$ are bounded, we have

$$\left\|\hat{V}_D-V_D\right\|_S=\left\|\hat{J}_L^{-1}\hat{V}_m\hat{J}_L^{-1}-V_D\right\|_S=o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)\tag{B.23}$$

where we have used the fact that

$$\left\|\hat{J}_L^{-1} - J^{-1}\right\|_S = \left\|J^{-1}(J - \hat{J}_L)\hat{J}_L^{-1}\right\|_S \leq \left\|\hat{J}_L^{-1}\right\|_S \left\|J^{-1}\right\|_S \left\|\hat{J}_L - J\right\|_S = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right). \quad \text{(B.24)}$$

Therefore, by definition of $\hat{V}_D$, equation (B.23) combined with Weyl's inequality and the fact that V_D has eigenvalues which are bounded and bounded form zero implies the eigenvalues of $\hat{V}_D$ are bounded and bounded from zero with probability approaching one. This establishes that $\|\hat{J}_L^{-1}\|_S$, $\|J^{-1}\|_S$, $\|\hat{V}_D\|_S$ and $\|V_D\|_S$ are bounded. Plugging these bounds and the bounds in equations (B.23) and (B.24) into equation (B.18) gives the desired result. □

Lemma B.6. ***Form of*** $\frac{d^2 Q_n(\theta)}{d\theta d\theta'}$ *Let Assumptions (B.1) - (B.5) hold. If* $\|\theta_1 - \theta_0\| = O_p\left(\sqrt{\frac{L}{K_L}}\right)$ *uniformly over* $\mathcal{P}$ *then*

$$\left\|\frac{d^2 Q_L(\theta_1)}{d\theta d\theta'} - \frac{d^2 Q_L(\theta_0)}{d\theta d\theta'}\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)$$

uniformly over $P \in \mathcal{P}$. *Moreover,*

$$\left\|\frac{d^2 Q_L(\theta_0)}{d\theta d\theta'} - \left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_0)}{d\theta'}\right)\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l, \theta_0)}{d\theta'}\right)'\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right).$$

Proof. Recall, $Q_L(\theta) = m_L(\theta)' m_L(\theta)$ where

$$m_L(\theta) = \frac{1}{L}\sum_{l=1}^{L} m(\boldsymbol{b}_l, \theta)$$

and where for any $\boldsymbol{b} \in \mathbb{R}^p$

$$m(\boldsymbol{b}, \theta) = \sum_{i=1}^{p}\left(\frac{d\log(g(b_i, \theta, p))}{d\theta} - \mu(\theta, p)\right)$$

where $\mu(\theta, p)$ is defined in equation (2.6). Therefore,

$$\frac{d^2 Q_L(\theta)}{d\theta d\theta'} = \frac{dm_L(\theta)}{d\theta}\frac{dm_L(\theta)}{d\theta'} + A_L(\theta)$$

where

$$[A_L(\theta)]_{ij} = \frac{d^2 m_L(\theta)}{d\theta_i d\theta_j} m_L(\theta).$$

For any $p \in \{2, ..., \overline{p}\}$ and any $b \in [\underline{v}, \overline{b}(\theta, p)]$ we have

$$\frac{d \log(g(b, \theta, p))}{d\theta} = (\psi_1(v, \theta, p) + \psi_2(v, \theta, p) + \psi_3(v, \theta, p))|_{v=v(b,\theta,p)}$$

where $v(b, \theta, p)$ is the inverse of $s(\cdot, \theta, b, p)$ in lemma (B.20) and $\{\phi_j\}_{j=1}^3$ are defined in lemma (B.21). By a mean-value expansion

$$m_L(\theta_1) = m_L(\theta_0) + \frac{dm_L(\theta^+)}{d\theta'}(\theta_1 - \theta_0) \tag{B.25}$$

where $\|\theta^+ - \theta_0\| \leq \|\theta_1 - \theta_0\|$. Further, by a coordinate-wise mean-value expansion

$$\left[\frac{dm_L(\theta^+)}{d\theta'} - \frac{dm_L(\theta_0)}{d\theta'}\right]_{ij} = (\theta^+ - \theta_0)' \frac{d^2 m_L(\tilde{\theta})}{d\theta_i d\theta_j}$$

where $\tilde{\theta}$ is an intermediate value (depending on i and j). Using this expression and lemma yields (B.22)

$$\begin{aligned}
\left\|\frac{dm_L(\theta^+)}{d\theta'} - \frac{dm_L(\theta_0)}{d\theta'}\right\|_S^2 &\leq \left\|\frac{dm_L(\theta^+)}{d\theta'} - \frac{dm_L(\theta_0)}{d\theta'}\right\|_F^2 \\
&= \sum_{i=1}^{K_L} \sum_{j=1}^{K_L} \left|\frac{d^2 m_L(\tilde{\theta})}{d\theta_j d\theta_i}(\theta^+ - \theta_0)\right|^2 \\
&\leq \sum_{i=1}^{K_L} \sum_{j=1}^{K_L} \left\|\frac{d^2 m_L(\tilde{\theta})}{d\theta_j d\theta_i}\right\|^2 \|\theta^+ - \theta_0\|^2 \\
&= \|\theta^+ - \theta_0\|^2 \sum_{i=1}^{K_L} \left\|\frac{d^2 m_L(\tilde{\theta})}{d\theta' d\theta_i}\right\|_F^2 \\
&\leq \|\theta^+ - \theta_0\|^2 C^2 \zeta_L^4 K_L \frac{1}{L} \sum_{l=1}^{L} \sum_{i=1}^{p_l} \frac{1}{(v(b_{il}, \tilde{\theta}, p_l) - \underline{v})}
\end{aligned}$$

for some $C < \infty$. As $\|\tilde{\theta} - \theta_0\| \leq \|\theta^+ - \theta_0\| \leq \|\theta_1 - \theta_0\|$, lemma (B.15) and the previous equation implies

$$\left\|\frac{dm_L(\theta^+)}{d\theta'} - \frac{dm_L(\theta_0)}{d\theta'}\right\|_F = O_p(1)\|\theta_1 - \theta_0\|\zeta_L^2 \sqrt{K_L} = o_p(\varepsilon_L) \tag{B.26}$$

were the last line uses Assumption (B.4) and $\|\theta_1 - \theta_0\| = O_p\left(\sqrt{\frac{K_L}{L}}\right)$. Therefore, by the reverse triangle inequality

$$\left\|\frac{dm_L(\theta^+)}{d\theta'}\right\|_S = \left\|\frac{dm_L(\theta_0)}{d\theta'}\right\|_S + o_p(\varepsilon_L). \tag{B.27}$$

Since $\left\|\frac{dm_L(\theta_0)}{d\theta'}\right\|_S$ is bounded with probability approaching one by lemma (B.9) combined with Assumption (B.5), $\left\|\frac{dm_L(\theta^+)}{d\theta'}\right\|_S$ is also bounded with probability approaching one. Combining this with equation (B.25) yields

$$\big|\|m_L(\theta_1)\| - \|m_L(\theta_0)\|\big| \leq C\|\theta_1 - \theta_0\|$$

with probability approaching one. When combined with lemma (B.9), we get

$$\|m_L(\theta_1)\| \leq C\|\theta_1 - \theta_0\| + \|m_L(\theta_0)\| = O_p\Big(\sqrt{\frac{K_L}{L}}\Big) + \big\|\mathbb{E}_0\big(m_L(\theta_0)\big)\big\| + o_p(\varepsilon_L)$$

Then by lemma (B.8),

$$\begin{aligned}\sqrt{L}\|m_L(\theta_1)\| &\leq C\|\sqrt{L}\,(\theta_1 - \theta_0)\| + \sqrt{L}\|m_L(\theta_0)\| \\ &\leq C\|\sqrt{L}\,(\theta_1 - \theta_0)\| + \|S_L\| + o_p(\varepsilon_L) \\ &= O_p(\sqrt{K_L}) \end{aligned} \tag{B.28}$$

where S_L is defined in Assumption (B.8) and the last equality follows by the premise $\|\theta_1-\theta_0\| = O_p\Big(\sqrt{\frac{K_L}{L}}\Big)$ and $\|S_L\| = O_p(\sqrt{K_L})$ which follows by Markov's inequality and Assumption (B.5). Therefore, notice

$$\begin{aligned}\Big\|A(\theta_1)\Big\| = \sqrt{\sum_{i=1}^{K_L}\sum_{j=1}^{K_L}\left|\frac{d^2m(\theta_1)}{d\theta_i d\theta_j}m(\theta_1)\right|^2} &\leq \Big\|m(\theta_1)\Big\|\sqrt{\sum_{j=1}^{K_L}\left\|\frac{d^2m(\theta_1)}{d\theta' d\theta_j}\right\|^2} \\ &\leq C\Big\|m(\theta_1)\Big\|\sqrt{K_L}\zeta_L^2 O_p(1).\end{aligned}$$

Hence, equation (B.28) shows $\|A(\theta_1)\| = o_p(\varepsilon_L K_L^{-\frac{1}{2}})$. This immediately establishes the second claim of the lemma when combined with Assumption (B.5). Moreover,

$$\left\|\frac{d^2Q_L(\theta_1)}{d\theta d\theta'} - \frac{d^2Q_L(\theta_0)}{d\theta d\theta'}\right\| \leq \left\|\frac{dm(\theta_1)}{d\theta}\frac{dm(\theta_1)}{d\theta'} - \frac{dm(\theta_0)}{d\theta}\frac{dm(\theta_0)}{d\theta'}\right\| + \Big\|A(\theta_0)\Big\| + \Big\|A(\theta_1)\Big\|.$$

Notice

$$
\begin{aligned}
\left\|\frac{dm(\theta_1)}{d\theta}\frac{dm(\theta_1)}{d\theta'}-\frac{dm(\theta_0)}{d\theta}\frac{dm(\theta_0)}{d\theta'}\right\| &\leq \left\|\frac{dm(\theta_1)}{d\theta}\right\|_S\left\|\frac{dm(\theta_0)}{d\theta}-\frac{dm(\theta_1)}{d\theta}\right\| \\
&+\left\|\frac{dm(\theta_0)}{d\theta}\right\|_S\left\|\frac{dm(\theta_0)}{d\theta}-\frac{dm(\theta_1)}{d\theta}\right\| \\
&\leq 2C\zeta_L^2\sqrt{K_L}\|\theta_1-\theta_0\| \\
&= 2C\frac{\zeta_L^2 K_L}{\sqrt{L}}\frac{1}{\sqrt{K_L}}\|\sqrt{L}\,(\theta_1-\theta_0)\| \\
&= o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)
\end{aligned}
$$

where the second inequality uses the fact that $\frac{dm(\theta_1)}{d\theta'}$ has (absolute) singular values which are bounded and bounded away from zero[2] and the third inequality uses Assumption (B.4) and the premise that $\|\sqrt{L}\,(\theta_1-\theta_0)\| = O_p(\sqrt{K_L})$.

□

Lemma B.7. ***Bounded Eigenvalues*** *Assumptions (B.3) - (B.5) imply the eigenvalues of J_L and Σ_L are bounded and bounded away from zero.*

Proof. By the second statement in lemma (B.6), the eigenvalues of J_n within an an arbitrarily small neighborhood of the eigenvalues of $\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta'}\right)\left(\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta'}\right)'$ by Weyl's inequality. Moreover, by lemma (B.9) and Assumption (2.11)

$$
\left\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta'}-\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)\right\| = o_p(\varepsilon_L).
$$

Hence, by Weyl's theorem again, the absolute value of the eigenvalues of $\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_0)}{d\theta'}$ are bounded and bounded away from zero as Assumption (B.5) implies the absolute value of the eigenvalues of $\mathbb{E}_0\left(\frac{dm(\boldsymbol{b},\theta_0)}{d\theta'}\right)$ are bounded and and bounded from zero. Consequently, by properties of eigenvalues, the eigenvalues of J_n are bonded and bounded away from zero with

[2]This follows as equation (B.27) is valid with θ_1 in place of θ^+ combined with lemma (B.9) and Assumption (B.5).

probability approaching one. For the next part of the proof, note

$$\Sigma_L \equiv J_L^{-1} Var(m_L(\theta_0)) J_L.$$

Therefore $\|\Sigma_L\|_S$ is bounded as $\|\Sigma_L\|_S \leq \|J_L^{-1}\|_S^2 \|Var(m_L(\theta_0))\|_S$ which is bounded by Assumption (B.5) and the first statement of this lemma. Moreover, with probability approaching one Σ_L is invertible (as J_L is invertible with probability approaching one and $Var(m_L(\theta_0))$ is invertible by assumption). When Σ_L is invertible, note

$$\|\Sigma_L^{-1}\|_S \leq \|J_L(Var(m_L(\theta_0)))^{-1} J_L\|_S < \|J_L\|_S^2 \|Var(m_L(\theta_0)\|_S$$

so that $\|\Sigma_L^{-1}\|_S$ is bounded with probability approaching one. Hence, all eigenvalues of Σ_L are bounded and bounded away from zero.

$\square$

B.2.4 Auxiliary Lemmas

Lemma B.8. ***Coupling*** *Suppose Assumptions (B.1) - (B.4) hold. For any $p \in \{2, ..., \overline{p}\}$ and $\boldsymbol{b} \in [\underline{v}, \overline{b}_0(p)]^p$ define*

$$m(\boldsymbol{b}, \theta) = \sum_{i=1}^{p} \left(\frac{d \log(g(b_i, \theta, p))}{d\theta} - \mu(\theta, p) \right)$$

where $\mu(\theta, p)$ is defined in equation (2.6). There exists a sequence $\{V_L, S_L\}$ with

$$\left\| \frac{1}{\sqrt{L}} \sum_{l=1}^{L} m(\boldsymbol{b}_l, \theta_{0,L}) - S_L \right\| = o_p\left(\varepsilon_L\right)$$

such that $S_L \sim N(0, V_L)$ and $V_L = Var(m(\boldsymbol{b}, \theta_{0,L}))$ where $\theta_{0,L}$ is defined in Assumption (B.2).

Proof. Define $\xi_l = m(\boldsymbol{b}_l, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))$ for each $l \in \{1, ..., L\}$ where the expectation is taken over $\boldsymbol{b}$. By Assumption (B.1), $\{\xi_j\}_{l=1}^{L}$ is independent and identically distributed with

$\mathbb{E}_0(\xi_l) = 0$. By lemma (B.16), there exists a constant $B < \infty$ such that for any $p \in \{2, ..., \overline{p}\}$ and any $\boldsymbol{b} \in [\underline{v}, \overline{b}_0(p)]^p$,

$$\|m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\| \leq \|m(\boldsymbol{b}, \theta_{0,L})\| + \mathbb{E}(\|m(\boldsymbol{b}, \theta_{0,L})\|) \leq 2\overline{p}B\zeta_n. \tag{B.29}$$

Additionally, as the trace operator is linear and invariant to cyclic permutations,

$$\begin{aligned}
&\mathbb{E}_0\left(\|m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\|^2\right) \\
&\quad = \operatorname{tr}\Big(\mathbb{E}_0\Big[\big(m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\big)'\big(m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\big)\Big]\Big) \\
&\quad = \mathbb{E}_0\Big[\operatorname{tr}\Big(\big(m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\big)\big(m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\big)'\Big)\Big] \\
&\quad = \operatorname{tr}\Big(\mathbb{E}_0\Big[\big(m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\big)\big(m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\big)'\Big]\Big) \\
&\quad \leq \operatorname{tr}\Big(\mathbb{E}_0\Big[m(\boldsymbol{b}, \theta_{0,L})m(\boldsymbol{b}, \theta_{0,L})'\Big]\Big) \\
&\quad \leq BK_n
\end{aligned}$$

where the expectations are taken over $\boldsymbol{b}$ and the second to last inequality uses the fact that the trace is non-negative for positive semi-definite matrices and the fact that the trace is the sum of the singular values of a matrix combined with Assumption (B.5)). By Assumption (B.1) and the above inequalities,

$$\begin{aligned}
\Delta_L &\equiv \sum_{l=1}^{L} \mathbb{E}_0\left(\|\xi_l\|^3\right) \\
&= \frac{1}{\sqrt{L}} \mathbb{E}_0\left(\|m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\|^3\right) \\
&\leq \frac{1}{\sqrt{L}} 2\overline{p}B\zeta_n \mathbb{E}_0\left(\|m(\boldsymbol{b}, \theta_{0,L}) - \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\|^2\right) \\
&= \frac{1}{\sqrt{L}} 2\overline{p}B^2\zeta_n K_n.
\end{aligned} \tag{B.30}$$

Define $T_L = \sum_{l=1}^{L} \xi_l$ and note

$$T_L = \left(\frac{1}{\sqrt{L}} \sum_{l=1}^{L} \sum_{i=1}^{p_l} \left(\frac{d \log(g(b_{il}, \theta_{0,L}, p_l)}{d\theta} - \mu(\theta_{0,L}, p_l)\right)\right) - \sqrt{L}\, \mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L})). \tag{B.31}$$

By Yurinskii's coupling, for any $\delta > 0$, there exists a random variable S_L with $S_L \sim N(0, V(T_L))$ such that

$$P\left(\|S_L - T_L\| > 3\delta\varepsilon_L\right) \leq C_0 \Delta_L K_L \left(\delta\varepsilon_L\right)^{-3} \left(1 + \frac{\left|\log\left(\Delta_L K_L \left(\delta\varepsilon_L\right)^{-3}\right)\right|}{K_L}\right)$$

$$= C_0 \delta^{-3} \Delta_L K_L \varepsilon_L^{-3} + C_0 \delta^{-3} \Delta_L \varepsilon_L^{-3} |\log(\delta^{-3} \Delta_L K_n \varepsilon_L^{-3})| \qquad \text{(B.32)}$$

for a universal constant C_0. We now use equation (B.30) and the rate conditions in Assumption (B.4) to provide bounds for both terms in equation (B.32). For the first term, the bound in equation (B.30) and the rate condition of Assumption (B.4) implies

$$C_0 \delta^{-3} \Delta_L K_L \varepsilon_L^{-3} \leq 2\overline{p} C_0 B^2 \delta^{-3} \frac{\zeta_L K_L^2}{\varepsilon_L^3 \sqrt{L}} = o_p(1). \qquad \text{(B.33)}$$

For the second term, equation (B.30) and Assumption (B.4) imply $\Delta_L K_L \varepsilon_L^{-3} = o_p(1)$, so that

$$\left|\log(\delta^{-3} \Delta_L K_L \varepsilon_L^{-3})\right| = \log\left(\frac{\delta^3 \varepsilon_L^3}{\Delta_L K_L}\right) + o_p(1) \leq \frac{\delta^3 \varepsilon_L^3}{\Delta_L K_L} + o_p(1)$$

where the first equality follows from equation (B.33) and the inequality follows as $\log(x) \leq x$.[3] Therefore,

$$C_0 \delta^{-3} \Delta_L \varepsilon_L^{-3} \left|\log(\delta^{-3} \Delta_L K_L \varepsilon_L^{-3})\right| \leq C_0 \frac{1}{K_L} + o_p(1) = o_p(1).$$

This establishes the second inequality. Therefore, by Yurinskii's coupling,

$$P\left(\|S_L - T_L\| > 3\delta\varepsilon_n\right) = o_p(1) \Rightarrow \|S_L - T_L\| = o_p(\varepsilon_L).$$

[3] To see this, let $a_L \equiv \delta^{-3} \Delta_L K_L \varepsilon_L^{-3}$ and A_L be the event that a_L is less than 1. By equation (B.33), $P(A_L^C) = o(1)$. Therefore, for any $\kappa > 0$

$$P\left(\left|\,|\log(a_L)| - \log(a_L^{-1})\right| > \kappa\right) \leq P\left(\left|\,|\log(a_L)| - \log(a_L^{-1})\right| > \kappa\ ,\ A_L\right) + P(A_L^C) = P(A_L^C) = o(1)$$

where the first equality follows as the event A_L implies $|\log(a_L)| = -\log(a_L) = \log(a_L^{-1})$.

To establish the lemma, note by equation (B.31) and the Cauchy-Schwarz inequality

$$\begin{aligned}\Bigg\|\frac{1}{\sqrt{L}}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\Big(\frac{d\log(g(b_{il},\theta_{0,L},p_l)}{d\theta}-\mu(\theta_{0,L},p_l)\Big)-S_L\Bigg\| \\ &\leq \sqrt{L}\;\|\mathbb{E}_0\left(m(\boldsymbol{b},\theta_{0,L})\right)\|+\|T_L-S_L\| \\ &= o_p(\varepsilon_L)\end{aligned}$$

where the last equality follows from lemma (B.10) so that

$$\sqrt{L}\,\mathbb{E}_0\left(\|m(\boldsymbol{b},\theta_{0,L})\|\right)=O_p\left(\sqrt{L}K_L^{\frac{1}{2}-\gamma}\right)=o_p(\varepsilon_L)$$

where the second equality uses Assumption (B.4). Finally, as the variables are iid, we have

$$V(T_L)=LV(\xi_1)=V\left(m(\boldsymbol{b},\theta_{0,L})\right),$$

which establishes the desired result. □

Lemma B.9. ***Convergence of Sample Moments*** *Suppose Assumptions (B.3)-(B.5) hold. For any $p\in\{2,...,\overline{p}\}$ and $\boldsymbol{b}\in[\underline{v},\overline{b}_0(p)]^p$ let*

$$m(\boldsymbol{b},\theta_{0,L})=\sum_{i=1}^{p}\Big(\frac{d\log(g(b_i,\theta_{0,L},p)}{d\theta}-\mu(\theta_{0,L},p)\Big)$$

where $\theta_{0,L}$ is defined in Assumption (B.2). Then

$$\Big\|\frac{1}{L}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_{0,L})-\mathbb{E}_0\Big(m(\boldsymbol{b},\theta_{0,L})\Big)\Big\|=O_p\Big(\frac{\zeta_L}{\sqrt{L}}\Big)$$

and

$$\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L},p)}{d\theta}-\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|=O_p\Big(\frac{\zeta_L^2}{\sqrt{L}}\Big)$$

where the expectations are taken over ***b****.*

Proof. We first establish the second statement. By Assumption (B.1), we have

$$\mathbb{E}_0\Big(\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta}-\mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|^2\Big)\leq\frac{1}{L}\mathbb{E}_0\Big(\Big\|\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big\|^2\Big).\tag{B.34}$$

For any $p \in \{2, ..., \overline{p}\}$ and any $b \in [\underline{v}, \overline{b}_0(p)]$ notice lemmas (B.16) and (B.23) implies

$$\begin{aligned}\Big\|\frac{d^2 \log(g(b,\theta_{0,L},p)}{d\theta d\theta'}\Big\| &\leq \frac{\Big\|\frac{d^2 g(b,\theta_{0,L},p)}{d\theta d\theta'}\Big\|}{g(b,\theta_{0,L},p)} + \frac{\Big\|\frac{dg(b,\theta_{0,L},p)}{d\theta}\Big\|^2}{g(b,\theta_{0,L},p)^2} \\ &\leq B^2(1+B^2)\zeta_L^2\end{aligned}$$

and

$$\Big\|\frac{d\mu(\theta_{0,L},p)}{d\theta'}\Big\| \leq B\zeta_n^2.$$

Combining these inequalities with the definition of $m(\boldsymbol{b},\theta_{0,L})$ shows

$$\Big\|\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big\| \leq p(B^2(1+B^2)+B)\zeta_L^2.$$

Hence, for each $p \in \{2, ..., \overline{p}\}$,, we have

$$\mathbb{E}_0\Big(\Big\|\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta'}\Big\|^2 \;\Big|\; \boldsymbol{b} \in \mathbb{R}^p\Big) \leq p^2(B^2(1+B^2)+B)\zeta_n^4 \leq \overline{p}^2(B^2(1+B^2)+B^4)\zeta_L^4.$$

Therefore, by the law of iterated expectations, $\mathbb{E}_0\Big(\Big\|\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big\|^2\Big) = O_p(\zeta_L^4)$. Therefore, by equation (B.34) we have

$$\mathbb{E}_0\Big(\Big\|\frac{1}{L}\sum_{l=1}^{L}\frac{dm(\boldsymbol{b}_l,\theta_{0,L})}{d\theta} - \mathbb{E}_0\Big(\frac{dm(\boldsymbol{b},\theta_{0,L})}{d\theta}\Big)\Big\|^2\Big) = O_p\Big(\frac{\zeta_L^4}{L}\Big).$$

The desired statement then follows from an application of Markov's inequality. To establish the first statement, by Markov's inequality, it suffices to show

$$\mathbb{E}_0\left(\|m_L(\theta_0) - \mathbb{E}_0(m_L(\theta_0))\|\right) = O_p(\sqrt{K_L}).$$

To establish this, notice by the triangle inequality and lemma (B.10),

$$\begin{aligned}\mathbb{E}_0\left(\|m_L(\theta_0) - \mathbb{E}_0(m_L(\theta_0))\|\right) &\leq \mathbb{E}_0\Big(\Big\|m_L(\theta_0)\Big\|\Big) + \|\mathbb{E}_0(m_L(\theta_0))\| \\ &\leq \mathbb{E}_0\Big(\Big\|m_L(\theta_0)\Big\|\Big) + O_p\Big(K_L^{\frac{1}{2}-\gamma}\Big). \qquad \text{(B.35)}\end{aligned}$$

for some constant $C < \infty$. Next, Fix any $p \in \{2, ..., \overline{p}\}$ and $\boldsymbol{b} \in [\underline{v}, \overline{b}_0(p)]$, note

$$\begin{aligned}\|m(\boldsymbol{b}, \theta_0)\| &\leq \sum_{i=1}^{p_l} \left(\left\| \frac{d \log(g(b_i, \theta_0, p)}{d\theta} \right\| + \|\mu(\theta_0, p)\| \right) \\ &\leq C\sqrt{K_L} \sum_{i=1}^{p_l} \left(\frac{1}{(v(b_i, \theta_0, p) - \underline{v})^{\frac{1}{2}}} + 1 \right)\end{aligned}$$

where $v(b_i, \theta_0, p)$ is the unique value v such that $b_i = v - \frac{\int_{\underline{v}}^{v} F(t,\theta_0)^{p-1} dt}{F(v,\theta_0)^{p-1} dt}$. By lemma (B.13), and this expression

$$\begin{aligned}\|m(\boldsymbol{b}, \theta_0)\| &\leq \sum_{i=1}^{p_l} \left(\left\| \frac{d \log(g(b_i, \theta_0, p)}{d\theta} \right\| + \|\mu(\theta_0, p)\| \right) \\ &\leq C\sqrt{K_L} \sum_{i=1}^{p_l} \left(C \frac{1}{(v(b_i, p) - \underline{v})^{\frac{1}{2}}} + 1 \right)\end{aligned}$$

where $v(b_i, p)$ is the unique value of v such that $b_i = v - \frac{\int_{\underline{v}} F_0(t)^{p-1} dt}{F_0(v)^{p-1} dt}$. Therefore, by Assumption (B.1) for any $\tilde{p} \in \{2, ..., \overline{p}\}$ we have

$$\begin{aligned}\mathbb{E}_0 \left(\|m(\boldsymbol{b}, \theta_0)\| \mid \boldsymbol{b} \in \mathbb{R}^{\tilde{p}} \right) &\leq C\sqrt{K_L}\tilde{p} \left(C \int_{\underline{v}}^{v} \frac{1}{(v - \underline{v})^{\frac{1}{2}}} f_0(v) dv + 1 \right) \\ &= C\sqrt{K_L}\tilde{p} \left(2C^2 (\overline{v} - \underline{v})^{\frac{1}{2}} + 1 \right).\end{aligned}$$

Therefore by the law of iterated expectations and (possibly) re-defining the constant C, there exists $C < \infty$ with

$$\mathbb{E}_0 \left(\|m(\boldsymbol{b}, \theta_0)\| \right) \leq C\sqrt{K_L}. \tag{B.36}$$

Combining this with expression (B.35) gives

$$\mathbb{E}_0 \left(\|m_L(\theta_0) - \mathbb{E}_0(m_L(\theta_0))\| \right) \leq C\sqrt{K_L} + O_p(K_L^{\frac{1}{2} - \gamma}) = O_p(\sqrt{K_L})$$

where we have used the assumption that γ is greater than one. This establishes the final statement of the lemma. □

Lemma B.10. *For any* $p \in \{2, ..., \overline{p}\}$ *and* $\boldsymbol{b} \in [\underline{v}, \overline{b}_0(p)]^p$ *define*

$$m(\boldsymbol{b}, \theta) = \frac{1}{L}\sum_{l=1}^{L}\Big(\frac{dm(b_i, \theta, p)}{d\theta} - \mu(\theta, p)\Big).$$

If Assumptions (B.1) - (B.3) hold, then $\|\mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}))\| = O_p\left(K_L^{\frac{1}{2}-\gamma}\right)$ *where* $\theta_{0,L}$ *and* γ *are given in Assumption (B.2).*

Proof. Fix any $p \in \{2, ..., \overline{p}\}$ and let $B < \infty$ be the maximum of the constants appearing in lemmas (B.11)-(B.16) and Assumption (B.3). By Assumption (B.1), for each $\boldsymbol{b} \in \mathbb{R}^p$, $g(\boldsymbol{b}, \theta_{0,L}, p) = \prod_{i=1}^{p} g(b_i, \theta_{0,L}, p)$.[4] Therefore,

$$\|\mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}) \,|\, \boldsymbol{b} \in \mathbb{R}^p)\| = p\Big\| \int_{\underline{v}}^{\overline{b}_0(p)} \frac{d\log(g(b, \theta_{0,L}, p))}{d\theta} g_0(b, p) - \mu(\theta_{0,L}, p)\Big\|.$$

Using this expression and the definition of $\mu(\theta_{0,L}, p)$ and assumption that $\overline{b}_0(p) \leq \overline{b}(\theta_{0,L}, p)$, we have

$$\begin{aligned}
\|\mathbb{E}_0(m(\boldsymbol{b}, \theta_{0,L}) \,|\, \boldsymbol{b} \in \mathbb{R}^p)\| &= \Big\| \int_{\underline{v}}^{\overline{b}_0(p)} \frac{d\log(g(b, \theta_{0,L}, p))}{d\theta} g_0(b, p)db - \mu(\theta_{0,L}, p)\Big\| \\
&= \Big\| \int_{\underline{v}}^{\overline{b}_0(p)} \frac{d\log(g(b, \theta_{0,L}, p))}{d\theta} g_0(b, p)db - \int_{\underline{v}}^{\overline{b}(\theta_{0,L},p)} \frac{d\log(g(b, \theta_{0,L}, p))}{d\theta} g(b, \theta_{0,L}, p)db\Big\| \\
&\leq \int_{\underline{v}}^{\overline{b}_0(p)} \Big\|\frac{d\log(g(b, \theta_{0,L}, p))}{d\theta}\Big\| |g_0(b, p) - g(b, \theta_{0,L}, p)|db \\
&\quad + \int_{\overline{b}_0(p)}^{\overline{b}(\theta_{0,L},p)} \Big\|\frac{d\log(g(b, \theta_{0,L}, p))}{d\theta}\Big\| g(b, \theta_{0,L}, p)db \qquad \text{(B.37)}
\end{aligned}$$

where the inequality follows as $g_0(b, p) = 0$ for all $b > \overline{b}_0(p)$. To bound the first term,

$$\begin{aligned}
&\int_{\underline{v}}^{\overline{b}_0(p)} \Big\|\frac{d\log(g(b, \theta_{0,L}, p))}{d\theta}\Big\| |g_0(b, p) - g(b, \theta_{0,L}, p)|db \\
&\qquad \leq B^2 K_L^{\frac{1}{2}} K_L^{-\gamma} \int_{\underline{v}}^{\overline{b}(\theta_{0,L},p)} (s^{-1}(b, \theta_{0,L}, p) - \underline{v})^{-\frac{1}{2}} db \\
&\qquad \leq (\overline{p} - 1)(\overline{v} - \underline{v})^{\frac{1}{2}} B^{2\overline{p}+1} K_L^{\frac{1}{2}-\gamma} \qquad \text{(B.38)}
\end{aligned}$$

[4]This follows as $g(\boldsymbol{b}, \theta_{0,L}, p) = f_{\boldsymbol{v}}(s^{-1}(\boldsymbol{b}, \theta_{0,L}, p), \theta_{0,L}) = \prod_{i=1}^{p} f(s^{-1}(b_i, \theta_{0,L}, p), \theta_{0,L})$ where $f_{\boldsymbol{v}}(\cdot, \theta_{0,L})$ denotes the joint density of the vector $\boldsymbol{v} = (v_1, ..., v_p)$ and $s(\cdot, \theta_{0,L}, p)$ denotes the bidding function. The statement then follows from the definition of $g(b, \theta_{0,L}, p) = f(s^{-1}(b_i, \theta_{0,L}, p), \theta_{0,L})$.

where the first inequality follow from lemmas (B.11) and (B.16) where $s^{-1}(\cdot,\theta_{0,L},p)$ denotes the inverse bid function with parameters $(\theta_{0,L},p)$ and the second inequality follows as

$$\int_{\underline{v}}^{\bar{b}(\theta_{0,L},p)}(s^{-1}(b,\theta_{0,L},p)-\underline{v})^{-\frac{1}{2}}db=\int_{\underline{v}}^{\overline{v}}(t-\underline{v})^{-\frac{1}{2}}s'(t,\theta_{0,L},p)dt\leq(\overline{p}-1)B^{2\overline{p}-1}(\overline{v}-\underline{v})^{\frac{1}{2}}$$

where we used the change of variables $t=s^{-1}(b,\theta_{0,L},p)$ and

$$\begin{aligned}s'(t,\theta_{0,L},p)&=(p-1)f(t,\theta_{0,L})\frac{\int_{\underline{v}}^{t}F(x,\theta_{0,L})^{p-1}dx}{F(t,\theta_{0,L})^{p}}\\&\leq(\overline{p}-1)B\frac{B^{p-1}\int_{\underline{v}}^{t}(x-\underline{v})^{p-1}dx}{\frac{1}{B^{p-1}}(t-\underline{v})^{p}}\\&\leq\frac{1}{2}(\overline{p}-1)B^{2\overline{p}-1}.\end{aligned}$$

To bound the second term, notice by the mean value theorem and lemmas (B.16) and (B.12)

$$\begin{aligned}&\int_{\bar{b}_0(p)}^{\bar{b}(\theta_{0,L},p)}\left\|\frac{d\log(g(b,\theta_{0,L},p))}{d\theta}\right\|g(b,\theta_{0,L},p)db\\&\qquad=(\bar{b}(\theta_{0,L},p)-\bar{b}_0(p))\left\|\frac{d\log(g(\tilde{b},\theta_{0,L},p))}{d\theta}\right\|g(\tilde{b},\theta_{0,L},p)\\&\qquad\leq B^3K_L^{\frac{1}{2}-\gamma}(s^{-1}(\bar{b}_0(p),\theta_{0,L},p)-\underline{v})^{-\frac{1}{2}}\\&\qquad=O_p\Big(K_L^{\frac{1}{2}-\gamma}\Big)\end{aligned}\tag{B.39}$$

where the last equality follows as

$$s^{-1}(\bar{b}_0(p),\theta_{0,L},p)=\overline{v}+o_p(1)\Rightarrow(s^{-1}(\bar{b}_0(p),\theta_{0,L},p)-\underline{v})^{-\frac{1}{2}}=O_p(1)$$

Replacing terms in equation (B.37) with the bounds in equations (B.38) and (B.39) gives

$$\|\mathbb{E}_0(m(\boldsymbol{b},\theta_{0,L})\,|\,\boldsymbol{b}\in\mathbb{R}^p)\|=O_p\Big(K_L^{\frac{1}{2}-\gamma}\Big).$$

It then follows by the law of iterated expectations to get

$$\|\mathbb{E}_0(m(\boldsymbol{b},\theta_{0,L}))\|=O_p\Big(K_L^{\frac{1}{2}-\gamma}\Big).$$

□

Lemma B.11. *Under Assumptions (B.2) and (B.3), there exists a finite constant $C < \infty$ such that*

$$\sup_{b \in [\underline{v}, \bar{b}(\theta_{0,L},p)} |g_0(b,p) - g(b,\theta_{0,L},p)| \leq CK_L^{-\gamma}.$$

for any $p \in \{2, ..., \overline{p}\}$ where γ is defined in Assumption (B.2).

Proof. Let $B < \infty$ be the maximum of the constants defined in Assumptions (B.2), (B.3) and lemma (B.12). Let $p \in \{2, ..., \overline{p}\}$ and $b \in [\underline{v}, \bar{b}_0(p)]$, be arbitrary. Let v_L and v_0 denote the (respective) zeros of the functions

$$s(v,\theta,p) = (v-b) - \frac{\int_{\underline{v}}^{v} F(t,\theta)^{p-1}dt}{F(v,\theta)^{p-1}}$$
$$s_0(v,p) = (v-b) - \frac{\int_{\underline{v}}^{v} F_0(t)^{p-1}dt}{F_0(v)^{p-1}}.$$

Notice,

$$(p-1)\,|g_0(b,p) - g(b,\theta_{0,L},p)| = \left| \frac{F_0(v_0)^p}{\int_{\underline{v}}^{v_0} F_0(t)^{p-1}dt} - \frac{F(v_L,\theta_{0,L})^p}{\int_{\underline{v}}^{v_L} F(t,\theta_{0,L})^{p-1}dt} \right|$$

$$= \left| \frac{F_0(v_0)^p}{\int_{\underline{v}}^{v_0} F_0(t)^{p-1}dt} - \frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} \right| + \left| \frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} - \frac{F(v_n,\theta_{0,L})^p}{\int_{\underline{v}}^{v_L} F(t,\theta_{0,L})^{p-1}dt} \right| \tag{B.40}$$

For the first term, by the Cauchy-Schwarz and triangle inequalities, we have

$$\left| \frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} - \frac{F_0(v_0)^p}{\int_{\underline{v}}^{v_0} F_0(t)^{p-1}dt} \right| \leq |v_L - v_0| \left(\frac{pF_0(\tilde{v})f_0(\tilde{v})}{\int_{\underline{v}}^{\tilde{v}} F_0(t)^{p-1}dt} + \frac{F_0(\tilde{v})^{2p-1}}{\left(\int_{\underline{v}}^{\tilde{v}} F_0(t)^{p-1}dt\right)^2} \right)$$
$$\leq p^2 B^{2p-1}(1 + B^{2p-1}) \frac{|v_L - v_0|}{\tilde{v} - \underline{v}}$$

where the second inequality uses the inequality $\frac{1}{B} \leq \frac{F_0(v)}{v - \underline{v}} \leq B$ and $f_0(v) < B$ along with some algebra. If $v_L < v_0$, then $v_L - \underline{v} < \tilde{v} - \underline{v}$, so by the second inequality of lemma (B.13) and the previous display,

$$\left| \frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} - \frac{F_0(v_0)^p}{\int_{\underline{v}}^{v_0} F_0(t)^{p-1}dt} \right| \leq p^2 B^{2p-1}(1 + B^{2p-1}) \frac{|v_L - v_0|}{v_L - \underline{v}}$$
$$\leq p^2 B^{2p}(1 + B^{2p-1}) K_L^{-\gamma}.$$

If $v_0 < v_L$, then $v_0 - \underline{v} < \tilde{v} - \underline{v}$. Therefore, the first inequality of lemma (B.13) gives

$$\begin{aligned}\left|\frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} - \frac{F_0(v_0)^p}{\int_{\underline{v}}^{v_0} F_0(t)^{p-1}dt}\right| &\leq p^2 B^{2p-1}(1+B^{2p-1})\frac{|v_L - v_0|}{v_0 - \underline{v}} \\ &\leq p^2 B^{2p}(1+B^{2p-1})K_L^{-\gamma}.\end{aligned}$$

Lastly, if $v_L = v_0$, then the inequality follows trivially. In all cases, we have

$$\left|\frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} - \frac{F_0(v_0)^p}{\int_{\underline{v}}^{v_0} F_0(t)^{p-1}dt}\right| \leq p^2 B^{2p}(1+B^{2p-1})K_L^{-\gamma}.$$

To bound the second term in equation (B.40), note

$$\begin{aligned}\sup_v |f_0(v) - f(v,\theta_{0,L})| \leq BK_L^{-\gamma} &\Rightarrow |F_0(v) - F(v,\theta_{0,L})| \leq BK_L^{-\gamma}(v - \underline{v}) \\ &\Rightarrow |F_0(v)^p - F(v,\theta_{0,L})^p| \leq B^p K_L^{-\gamma}(v-\underline{v})^p\end{aligned}$$

where the last implication uses the difference of powers formula.[5] Therefore, by the triangle inequality

$$\begin{aligned}\left|\frac{F_0(v_L)^p}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} - \frac{F(v_L,\theta_{0,L})^p}{\int_{\underline{v}}^{v_L} F(t,\theta_{0,L})^{p-1}dt}\right| &\leq \frac{|F_0(v_L)^p - F(v_L,\theta_{0,L})^p|}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt} \\ &\quad + \frac{F(v_L,\theta_{0,L})^p \int_{\underline{v}}^{v_L} |F_0(t)^{p-1} - F(t,\theta_{0,L})^{p-1}|dt}{\int_{\underline{v}}^{v_L} F_0(t)^{p-1}dt \int_{\underline{v}}^{v_L} F(t,\theta_{0,L})^{p-1}dt} \\ &\leq pB^{2p-1}\left(1 + pB^{2p-2}\right)K_L^{-\gamma},\end{aligned}$$

which establishes the bound for the second term. Therefore, if we let

$$C = 2\max\left\{p^2B^{2p}(1+B^{2p-1}),\, pB^{2p-1}(1+pB^{2p-2})\right\}_{p=2}^{\overline{p}},$$

the previous two bounds and equation (B.40) establishes the lemma. □

[5]Specifically, by the difference of powers identity

$$|F_0(v)^p - F(v,\theta_{0,L})^p| = |F_0(v) - F(v,\theta_{0,L})|\sum_{i=0}^{p-1} F_0(v)^i F(v,\theta_{0,L})^{p-1-i} \leq B^p K_L^{-\gamma}(v-\underline{v})^p$$

where the last inequality uses $F_0(v) \leq B(v-\underline{v})$, $F(v) \leq B(v-\underline{v})$ and the premise $|F_0(v) - F(v,\theta_{0,L})| \leq BK_n^{-\gamma}(v-\underline{v})$.

Lemma B.12. *Under Assumptions (B.2) and (B.3), there exists $C < \infty$ with*

$$|\overline{b}(\theta_{0,L}, p) - \overline{b}_0(p)| \leq CK_L^{-\gamma},$$

for all $p \in \{2, ..., \overline{p}\}$ where

$$\overline{b}_0(p) \equiv \overline{v} - \int_{\underline{v}}^{\overline{v}} F_0(t)^{p-1} dt \quad \text{and} \quad \overline{b}(\theta_{0,L}, p) \equiv \overline{v} - \int_{\underline{v}}^{\overline{v}} F(t, \theta_{0,L})^{p-1} dt$$

and γ is defined in Assumption (B.2).

Proof. Let B be the maximum of the constants defined in Assumptions (B.2) and (B.3), and notice

$$\sup_{v \in [\underline{v}, \overline{v}]} |f_0(v) - f(v, \theta_{0,L})| \leq BK_L^{-\gamma} \Rightarrow |F_0(v) - F(v, \theta_{0,L})| \leq BK_L^{-\gamma}(v - \underline{v}).$$

Additionally, $f_0(v), f(v, \theta_{0,L}) \leq B$ for all $v \in [\underline{v}, \overline{v}]$ implies $F_0(v)^p, F(v, \theta_{0,L})^p \leq B^p(v - \underline{v})^p$ for all $p \in \{2, ..., \overline{p}\}$. By the difference of powers formula, and the above inequalities we have

$$\begin{aligned} |F_0(v)^{p-1} - F(v, \theta_{0,L})^{p-1}| &= |F_0(v) - F(v, \theta_{0,L})| \sum_{i=0}^{p-2} F_0(v)^i F(v, \theta_{0,L})^{p-2-i} \\ &\leq B^{p-1} K_L^{-\gamma} (v - \underline{v})^{p-1} \end{aligned}$$

Therefore,

$$\begin{aligned} |\overline{b}(\theta_{0,L}, p) - \overline{b}_0(p)| &\leq \int_{\underline{v}}^{\overline{v}} \left| F(t, \theta_{0,L})^{p-1} - F_0(t)^{p-1} \right| \\ &\leq B^{p-1} K_L^{-\gamma} \int_{\underline{v}}^{\overline{v}} (t - \underline{v})^{p-1} dt \\ &= \frac{(\overline{v} - \underline{v})^p}{p} B^{p-1} K_L^{-\gamma} \end{aligned}$$

Therefore, letting $C = \max\{\frac{1}{p}(\overline{v} - \underline{v})^p B^{p-1}\}_{p=2}^{\overline{p}}$ gives the result. □

Lemma B.13. *Under Assumptions (B.3) and (B.2), there exists a finite constant $C < \infty$ such that for any $p \in \{2, ..., \overline{p}\}$ and any $b \in [\underline{v}, \overline{b}_0(p)]$*

$$\frac{|v(b, \theta_{0,L}, p) - v_0(b, p)|}{v_0(b, p) - \underline{v}} \leq CK_L^{-\gamma}$$

$$\frac{|v(b, \theta_{0,L}, p) - v_0(b, p)|}{v(b, \theta_{0,L}, p) - \underline{v}} \leq CK_L^{-\gamma}$$

where $v(b,\theta_{0,L},p)$ *and* $v_0(b,p)$ *are the (unique) solutions to* $0 = s(v(b,\theta_{0,L},p),\theta_{0,L},p)$ *and* $0 = s_0(v_0(b,p),p)$ *where*

$$s(v,\theta_{0,L},p) \equiv (v-b) - \frac{\int_{\underline{v}}^{v} F(t,\theta_{0,L})^{p-1}dt}{F(v,\theta_{0,L})^{p-1}} \quad \textit{and} \quad s_0(v,p) \equiv (v-b) - \frac{\int_{\underline{v}}^{v} F_0(t)^{p-1}dt}{F_0(v)^{p-1}}.$$

Proof. As the second follows from a symmetric argument, we only proof the first inequality. Let B denote the maximum of the constants appearing in Assumptions (B.2) and (B.3). By assumption (B.2)

$$s'(v,\theta_{0,L},p) = (p-1)f(v,\theta_{0,L})\frac{\int_{\underline{v}}^{v} F(t,\theta_{0,L})^{p-1}dt}{F(v,\theta_{0,L})^{p}} > (p-1)B^{-2p} \tag{B.41}$$

with identical bounds for $s_0'(v,p)$. Therefore, by (possibly) re-defining B, there exists $B<\infty$ such that $\frac{1}{B} < s'(v,\theta_{0,L},p), s_0'(v,p) < B$. As $s_0(v_0(b,p),p) = 0$, we have

$$\begin{aligned}
&|s(v_0(b,p),\theta_{0,L},p)| = |s(v_0(b,p),\theta_{0,L},p) - s_0(v_0(b,p),p)| \\
&\leq \frac{\int_{\underline{v}}^{v_0(b,p)} |F(t,\theta_{0,L})^{p-1} - F_0(t)^{p-1}|dt}{F(v_0(b,p),\theta_{0,L})^{p-1}} \\
&\quad + \frac{\left|F_0(v_0(b,p))^{p-1} - F(v_0(b,p),\theta_{0,L},p)^{p-1}\right| \int_{\underline{v}}^{v_0(b,p)} F_0(t)^{p-1}dt}{F_0(v_0(b,p))^{p-1}F(v_0(b,p),\theta_{0,L})^{p-1}} \\
&\leq \frac{1}{p}B^{2p-2}(1+B^{2p-2})K_L^{-\gamma}(v_0(b,p)-\underline{v})
\end{aligned} \tag{B.42}$$

where the last line uses assumption (B.2) and the fact that

$$\begin{aligned}
\sup_{t\in[\underline{v},\overline{v}]} |f_0(t) - f(t,\theta_{0,L})| \leq BK_L^{-\gamma} &\Rightarrow |F(v,\theta_{0,L}) - F_0(v)| \leq BK_L^{-\gamma}(v-\underline{v}) \\
&\Rightarrow |F(v,\theta_{0,L})^p - F_0(v)^p| \leq B^pK_L^{-\gamma}(v-\underline{v})^p
\end{aligned}$$

where the last implication follows by the formula for the difference of powers and assumption (B.2).[6]

[6]By the difference of powers formula, we have

$$|F(v,\theta)^p - F_0(v)^p| = |F(v,\theta) - F_0(v)|\sum_{i=0}^{p-1} F(v,\theta)^{p-1-i}F_0(v)^i \leq D^p(v-\underline{v})^pK_L^{-\gamma}$$

where the inequality uses $F_0(v)^j \leq B^j(v-\underline{v})$ and $F(v,\theta_{0,L})^j \leq B^j(v-\underline{v})$

Now we proceed by cases. For the first case, if $s(v_0(b,p),\theta,p) = 0$ then $v_0(b,p) = v(b,\theta_{0,L},p)$ by the uniqueness of the zero of the function (see lemma (B.18)). For the second case, if $s(v_0(b,p),\theta_{0,L},p) > 0$, then $v_0(b,p) > v(b,\theta_{0,L},p)$ as $s(\cdot,\theta_{0,L},p)$ has a unique zero and is strictly increasing. Therefore, by equation (B.42),

$$\begin{aligned} BK_L^{-\gamma}(v_0(b,p) - \underline{v}) &> s(v_0(b,p),\theta_{0,L},p) - s(v(b,\theta_{0,L},p),\theta_{0,L},p) \\ &= \int_{v(b,\theta_{0,L},p)}^{v_0(b,p)} s'(t,\theta_{0,L},p)dt \\ &> \frac{1}{B}(v_0(b,p) - v(b,\theta_{0,L},p)) \end{aligned}$$

where we have used $s(v(b,\theta_{0,L},p),\theta,p) = 0$ and equation (B.41) to bound the integrand. Rearranging this gives

$$0 < \frac{v_0(b,p) - v(b,\theta_{0,L},p)}{v_0(b,p) - \underline{v}} \leq B^2 K_L^{-\gamma}.$$

In the last case, if $s(v_0(b,p),\theta_{0,L},p) < 0$ then $v_0(b,p) < v(b,\theta_{0,L},p)$ and a similar display as before shows

$$\begin{aligned} BK_L^{-\gamma}(v_0(b,p) - \underline{v}) &> s(v(b,\theta_{0,L},p),\theta_{0,L},p) - s(v_0(b,p),\theta_{0,L},p) \\ &= \int_{v_0(b,p)}^{v(b,\theta_{0,L},p)} s'(t,\theta_{0,L},p)dt \\ &> \frac{1}{B}(v(b,\theta,p) - v_0(b,p)) \end{aligned}$$

so that

$$0 < \frac{v(b,\theta,p) - v_0(b,p)}{v_0(b,p) - \underline{v}} \leq B^2 K_L^{-\gamma}.$$

In all cases, the desired inequality holds. □

Lemma B.14. *Under Assumptions (B.3) and (B.2), there exists a finite constant* $C < \infty$

such that for any $p \in \{2, ..., \overline{p}\}$*, any* $b \in [\underline{v}, \overline{b}_0(p)]$ *and any* $\theta_1, \theta_2 \in \hat{\Theta}_R$*,*

$$\frac{|v(b,\theta_1,p) - v(b,\theta_1,p)|}{v(b,\theta_1,p) - \underline{v}} \leq C\zeta_L\|\theta_1 - \theta_2\|$$
$$\frac{|v(b,\theta_1,p) - v(b,\theta_1,p)|}{v(b,\theta_1,p) - \underline{v}} \leq C\zeta_L\|\theta_1 - \theta_2\|$$

where $v_i = v(b, \theta_i, p)$ *are the (unique) solutions to* $0 = s(v_i, p), \theta_i, p)$ *where*

$$s(v,\theta,p) \equiv (v-b) - \frac{\int_{\underline{v}}^{v} F(t,\theta)^{p-1}dt}{F(v,\theta)^{p-1}}.$$

Proof. As the second follows from a symmetric argument, we only proof the first inequality. Let B denote the constant appearing in Assumption (B.3). By Assumption (B.3),

$$s'(v,\theta,p) = (p-1)f(v,\theta)\frac{\int_{\underline{v}}^{v} F(t,\theta)^{p-1}dt}{F(v,\theta)^{p}} > (p-1)B^{-2p} \tag{B.43}$$

with $s'(v,\theta,p) < (p-1) < B^{2p}$ derived similarly. Therefore, by (possibly) re-defining B, there exists $B < \infty$ such that $\frac{1}{B} < s'(v,\theta,p), s_0'(v,p) < B$. Furthermore, as $s(v_1,\theta_1,p) = 0$, we have

$$\begin{aligned}
|s(v_1,\theta_2,p)| &= |s(v_1,\theta_2,p) - s(v_1,\theta_1,p)| \\
&\leq \frac{1}{F(v_1,\theta_2)^{p-1}}\int_{\underline{v}}^{v_1} |F(t,\theta_1)^{p-1} - F(t,\theta_2)^{p-1}|dt \\
&\quad + \frac{\int_{\underline{v}}^{v_1} F(t,\theta_2)^{p-1}dt}{F(v_1,\theta_1)^{p-1}F(v_1,\theta_2)^{p-2}}|F(v_1,\theta_1)^{p-1} - F(v_1,\theta_2)^{p-1}| \\
&\leq B(1+B^2)\zeta_L\|\theta_1 - \theta_2\|
\end{aligned} \tag{B.44}$$

where the last line uses

$$|f(t,\theta_1) - f(t,\theta_2)| \leq |P(t)'(\theta_1 - \theta_2)||P(t)'(\theta_1 + \theta_2)|\phi(t) \leq 2B^2\zeta_L\|\theta_1 - \theta_2\|(v_1 - \underline{v}).$$

Therefore, by the binomial theorem and possibly re-defining the constant B, there exists a $0 < B < \infty$ such that for any v

$$|F(v,\theta_1)^{p-1} - F(v,\theta_2)^{p-1}| \leq B\zeta_L\|\theta_1 - \theta_2\|.$$

By re-defining the constant B, the previous displays show there exists $0 < B < \infty$ such that

$$|s(v_1, \theta_2, p)| \leq B\zeta_L \|\theta_1 - \theta_2\|(v_1 - \underline{v}) \tag{B.45}$$

and $\frac{1}{B} < s'(v, \theta, p) < B$ for all $v \in [\underline{v}, \bar{b}_0(p)]$.

Now we proceed by cases. For the first case, if $s(v_1, \theta_2, p) = 0$ then $v_1 = v_2$ and there is nothing to prove. For the second case, if $s(v_1, \theta_2, p) > 0$, then $v_1 > v_2$ as $s(\cdot, \theta_2, p)$ has a unique zero and is strictly increasing. Therefore, by equation (B.45),

$$\begin{aligned} B\zeta_L\|\theta_1 - \theta_2\|(v_1 - \underline{v}) &> s(v_1, \theta_2, p) - s(v_2, \theta_2, p) \\ &= \int_{v_2}^{v_1} s'(t, \theta_2, p)dt \\ &> \frac{1}{B}(v_1 - v_2) \end{aligned}$$

where we have used $s(v_2, \theta_2, p) = 0$ and the fact that $\frac{1}{B} < s'(v, \theta, p)$ to bound the integrand. Re-arranging this gives

$$0 < \frac{v_1 - v_2}{v_1 - \underline{v}} \leq B^2\zeta_L\|\theta_1 - \theta_2\|.$$

In the last case, if $s(v_1, \theta_2, p) < 0$ then $v_1 < v_2$ and a similar display as before shows

$$\begin{aligned} B\zeta_L\|\theta_1 - \theta_2\|(v_1 - \underline{v}) &> s(v_2, \theta_1, p) - s(v_1, \theta_1, p) \\ &= \int_{v_1}^{v_2} s'(t, \theta_1, p)dt \\ &> \frac{1}{B}(v_2 - v_1) \end{aligned}$$

so that

$$0 < \frac{v_2 - v_1}{v_1 - \underline{v}} \leq B^2\zeta_L\|\theta_1 - \theta_2\|.$$

In all cases, the desired inequality holds. □

Lemma B.15. *Suppose Assumptions (B.1)-(B.5) hold and let* $\|\theta_1 - \theta_0\| = O_p\left(\sqrt{\frac{K_L}{L}}\right)$ *uniformly over* $\mathcal{P}$*. Furthermore, for any* $p \in \{1, ..., \overline{p}\}$ *and any* $b \in [\underline{v}, \overline{b}_0(p)$ *let* $v(b, \theta_1, p)$ *be the unique value such that* $0 = s(v(b, \theta_1, p), b, \theta, p)$ *where*

$$s(v, b, \theta, p) = (v - b) - \frac{\int_{\underline{v}}^{v} F(t, \theta)^{p-1} dt}{F(v, \theta)^{p-1} dt}.$$

Then,

$$\frac{1}{L} \sum_{l=1}^{L} \sum_{i=1}^{p_l} \frac{1}{(v(b_{il}, \theta_1, p_l) - \underline{v})^{\frac{1}{2}}} = O_p(1)$$

uniformly over $\mathcal{P}$*.*

Proof. Let b_{il} denote an arbitrary element of an arbitrary observation from the data $\{\boldsymbol{b}_l\}_{l=1}^{L}$. Let $v_j = v_{\iota(i,l)}$ where $v_{\iota(i,l)}$ is defined in Assumption (B.1). By the triangle inequality and some algebra,

$$\begin{aligned}\left|\frac{1}{(v(b_{il}, \theta_1, p_l) - \underline{v})^{\frac{1}{2}}} - \frac{1}{(v_j - \underline{v})^{\frac{1}{2}}}\right| &\leq \sqrt{\frac{v_j - v(b_{il}, \theta_1, p)}{(v_j - \underline{v})(v(b_{il}, \theta_1, p) - \underline{v})}} \\ &\leq \sqrt{\frac{v_j - v(b_{il}, \theta_1, p)}{v(b_{il}, \theta_0, p_l) - \underline{v}}} \sqrt{\frac{v(b_{il}, \theta_1, p) - \underline{v}}{v(b_{il}, \theta_1, p) - \underline{v}}} \frac{1}{(v_j - \underline{v})^{\frac{1}{2}}} \\ &\quad + \sqrt{\frac{v(b_{il}, \theta_0, p_l) - v(b_{il}, \theta_1, p_l)}{v(b_{il}, \theta_1, p_l) - \underline{v}}} \frac{1}{(v_j - \underline{v})^{\frac{1}{2}}} \end{aligned} \tag{B.46}$$

where the first inequality uses the fact that $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|}$.[7] Let B denote the maximum of the constants appearing in lemmas (B.13) and (B.14). Then equation (B.46) implies

$$\left|\frac{1}{(v(b_{il}, \theta_1, p_l) - \underline{v})^{\frac{1}{2}}} - \frac{1}{(v_j - \underline{v})^{\frac{1}{2}}}\right| \leq \sqrt{B}\left(\sqrt{B} + 1\right) \zeta_L^{\frac{1}{2}} \left(1 + K_L^{-\frac{1}{2}\gamma}\right) \|\theta_1 - \theta_0\| \frac{1}{(v_j - \underline{v})^{\frac{1}{2}}} \tag{B.47}$$

Additionally, notice,

$$\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sqrt{v_j - \underline{v}}} = O_p(1) \tag{B.48}$$

[7]Assume without loss $x > y$ so that $\sqrt{x} = \sqrt{y + (x - y)} \leq \sqrt{y} + \sqrt{x - y}$ which implies $\sqrt{x} - \sqrt{y} \leq \sqrt{x - y}$.

uniformly over $P \in \mathcal{P}$ by Markov's inequality.[8] Therefore, equation (B.47) and Assumption (B.4) imply

$$\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\left|\frac{1}{(v(b_{il},\theta_1,p_l)-\underline{v})^{\frac{1}{2}}}-\frac{1}{(v_{\iota(i,l)}-\underline{v})^{\frac{1}{2}}}\right| = O_p(1)\frac{\zeta_L^{\frac{1}{2}}\sqrt{K_L}}{\sqrt{L}}\|\sqrt{L}\,(\theta_1-\theta_0)\|\frac{1}{\sqrt{K_L}}$$

$$= o_p(\varepsilon_L). \tag{B.49}$$

Therefore, by the triangle inequality and equations (B.48) and (B.49),

$$\left|\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\frac{1}{(v(b_{il},\theta_1,p_l)-\underline{v})^{\frac{1}{2}}}\right|$$

$$= \frac{1}{n}\sum_{j=1}^{n}\frac{1}{(v_j-\underline{v})^{\frac{1}{2}}}+\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\left|\frac{1}{(v(b_{il},\theta_1,p_l)-\underline{v})^{\frac{1}{2}}}-\frac{1}{(v_{\iota(i,l)}-\underline{v})^{\frac{1}{2}}}\right|$$

$$= O_p(1)$$

which establishes the desired result. □

Lemma B.16. *If Assumptions (B.3) and (B.1) hold there exists a constant* $0 < C < \infty$

$$g(b,\theta,p) \in \left[\frac{1}{B}, B\right]$$

$$\left\|\frac{dg(b,\theta,p)}{d\theta}\right\| \le B\min\left\{\sqrt{K_L}(v(b,\theta,p)-\underline{v})^{-\frac{1}{2}}\,,\,\zeta_L\right\}$$

$$\left\|\frac{d^2g(b,\theta,p)}{d\theta d\theta'}\right\| \le B\min\left\{\sqrt{K_L}\zeta_L(v(b,\theta,p)-\underline{v})^{-\frac{1}{2}}\,,\,\zeta_L^2\right\}$$

for all $\theta \in \Theta_{K_L}$, $p \in \{2,...,\overline{p}\}$ *and* $b \in (\underline{v},\overline{b}(\theta,p))$ *where* $v(b,\theta,p)$ *is the (unique) value such that* $b(v(b,\theta,p),\theta,p) = b$.

[8] The uniformity follows by Markov's inequality as

$$P\left(\frac{1}{n}\sum_{j=1}^{n}(v_j-\underline{v})^{-\frac{1}{2}} > a\right) \le \frac{1}{n}\frac{\mathbb{E}_0\left((v-\underline{v})^{-\frac{1}{2}}\right)}{a}$$

and uniformly over $\mathcal{P}$ we have

$$\mathbb{E}_0\left((v-\underline{v})^{-\frac{1}{2}}\right) = \int_{\underline{v}}^{\overline{v}}\frac{1}{(v-\underline{v})^{\frac{1}{2}}}f_0(v)dv \le C\int_{\underline{v}}^{\overline{v}}(v-\underline{v})^{-\frac{1}{2}}dv = 2C(\overline{v}-\underline{v})^{\frac{1}{2}} < \infty.$$

Proof. Let B denote the maximum of the constants appearing in lemmas (B.17) - (B.18). Fix $\theta \in \Theta_{K_L}$ and define

$$h(v,\theta,p) = \frac{F(v,\theta)^p}{\int_{\underline{v}}^v F(t,\theta)^{p-1}dt}$$

for any $v \in [\underline{v}, \overline{v}]$. Furthermore, for any $b \in (\underline{v}, \overline{b}(\theta,p))$. let $v(b,\theta,p)$ denote the unique value such that $b = b(v(b,\theta,p),\theta,p)$ (see lemma (B.18) for the existence and uniqueness claim). Using standard arguments, we can show

$$g(b,\theta,p) = h(v(b,\theta,p),\theta,p)$$

for any $b \in [\underline{v}, \overline{b}(\theta,p)]$. Notice, by lemma (B.17), $g(b,\theta,p) \leq B$. Moreover, by the bounds in lemma (B.19),

$$h(v,\theta,p) = \frac{F(v,\theta)^p}{\int_{\underline{v}}^v F(t,\theta)^{p-1}dt} > \frac{\frac{(v-\underline{v})^p}{B^p}}{\int_{\underline{v}}^v B^{p-1}(t-\underline{v})^{p-1}dt} = pB^{-2p}.$$

Additionally, by the chain rule and lemma (B.18)

$$\begin{aligned} \frac{dg(b,\theta,p)}{d\theta} &= \left(h_v(v(b,\theta,p),\theta,p)\frac{dv(b,\theta,p)}{d\theta} + h_\theta(v(b,\theta,p),\theta,p) \right) \\ &\equiv \left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p) \right)\big|_{v=v(b,\theta,p)} \end{aligned} \tag{B.50}$$

where

$$\psi(v,\theta,p) \equiv \frac{F(v,\theta)}{f(v,\theta)} \frac{\int_{\underline{v}}^v F(t,\theta)^{p-2}\frac{\partial F(t,\theta)}{\partial\theta}dt}{\int_{\underline{v}}^v F(t,\theta)^{p-1}dt} - \frac{\frac{\partial F(v,\theta)}{\partial\theta}}{f(v,\theta)}.$$

By the bounds in lemmas (B.17) and (B.18) combined with the triangle inequality,

$$\left\| \frac{dg(b,\theta,p)}{d\theta} \right\| \leq \left(B^3 + B^2 + B\right) \min\left\{ \sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}}, \zeta_L \right\}\Big|_{v=v(b,\theta,p)}.$$

Moreover, by the chain-rule and the above derivations,

$$\begin{aligned} \frac{d^2 g(b,\theta,p)}{d\theta d\theta'} &= \frac{\partial\left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p)\right)}{\partial v}\frac{dv(b,\theta,p)}{d\theta'} \\ &+ \frac{\partial\left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p)\right)}{\partial\theta'}\Bigg|_{v=v(b,\theta,p)} \end{aligned} \tag{B.51}$$

where

$$\frac{\partial \left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p)\right)}{\partial v}$$
$$= h_v(v,\theta,p)\psi_v(v,\theta,p) + h_{vv}(v,\theta,p)\psi(v,\theta,p) + h_{v\theta}(v,\theta,p)$$

and

$$\frac{\partial \left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p)\right)}{\partial \theta'}$$
$$= \psi(v,\theta,p)h_{v\theta}(v,\theta,p)' + h_v(v,\theta,p)\psi_\theta(v,\theta,p) + h_{\theta\theta}(v,\theta,p).$$

By lemmas (B.17) and (B.18),

$$\left\| \frac{\partial \left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p)\right)}{\partial v} \right\|$$
$$\leq |h_v(v,\theta,p)||\psi_v(v,\theta,p)| + |h_{vv}(v,\theta,p)||\psi(v,\theta,p)| + \|h_{v\theta}(v,\theta,p)\|$$
$$\leq B(2B+1)\zeta_L(v-\underline{v})^{-1}$$

and

$$\left\| \frac{\partial \left(h_v(v,\theta,p)\psi(v,\theta,p) + h_\theta(v,\theta,p)\right)}{\partial \theta'} \right\|$$
$$\leq \|\psi(v,\theta,p)\|\|h_\theta(v,\theta,p)\| + |h_v(v,\theta,p)|\|\psi_\theta(v,\theta,p)\| + \|h_{\theta\theta}(v,\theta,p)\|$$
$$\leq B(2B+1)\sqrt{K_L}\zeta_L(v-\underline{v})^{-\frac{1}{2}}.$$

Therefore, by lemma (B.18) and equation (B.51),

$$\left\| \frac{\partial^2 g(b,\theta,p)}{\partial\theta\partial\theta'} \right\| \leq B(2B+1)(B+1)\sqrt{K_L}\zeta_L(v(b,\theta,p)-\underline{v})^{-\frac{1}{2}}.$$

By the previous inequalities, taking

$$C = \max\left\{\max\left\{B\,,\ \frac{1}{p}B^{2p}\,,\ (B^3+B^2+B)\,,\ B(2B+1)(B+1)\right\}\right\}_{p=2}^{\overline{p}}$$

establishes the lemma. □

Lemma B.17. *Let $\theta \in \Theta_{K_L}$ and $v \in (\underline{v}, \overline{v})$, define*

$$h(v, \theta, p) = \frac{F(v, \theta)^p}{\int_{\underline{v}}^{v} F(v, \theta)^{p-1} dt}$$

If Assumption (B.3) holds there exists a constant $0 < C < \infty$

$$
\begin{aligned}
|h(v, \theta, p)| &\leq C \\
|h_v(b, \theta, p)| &\leq C(v - \underline{v})^{-1} \\
\|h_\theta(v, \theta, p)\| &\leq C \min\left\{ \sqrt{K_L}(v - \underline{v})^{-\frac{1}{2}} \,,\, \zeta_L \right\} \\
|h_{vv}(v, \theta, p)| &\leq C(v - \underline{v})^{-2} \\
\|h_{\theta v}(v, \theta, p)\| &\leq C\zeta_L(v - \underline{v})^{-1} \\
\|h_{\theta\theta}(\theta, \theta, p)\| &\leq C \min\left\{ \sqrt{K_L}\zeta_L(v - \underline{v})^{-\frac{1}{2}} \,,\, \zeta_L^2 \right\}
\end{aligned}
$$

Proof. For brevity, we only include the proof of the first, third and sixth inequalities. The proofs of the remaining inequalities follow similar arguments. Before establishing the lemma, we establish several useful inequalities. Let B denote the maximum of the constants appearing in Assumption (B.3) First, by Assumption (B.3), $\frac{1}{B} < f(v, \theta) < B$ implies

$$\frac{1}{B}(v - \underline{v}) < F(v, \theta) < B(v - \underline{v}). \tag{B.52}$$

Further, notice as $f(v, \theta) = (P(v)'\theta)^2 \phi(v)$, we have

$$\left\| \frac{\partial f(v, \theta)}{\partial \theta} \right\|^2 \leq 4\|P(v)\|^2 (P(v)'\theta)^2 \phi(v)^2 = 4\|P(v)\|^2 f(v, \theta)\phi(v) \leq 4B\|P(v)\|^2 \phi(v).$$

Therefore, by the definition of ζ_L and the previous inequality, we have

$$\left\| \frac{\partial F(v, \theta)}{\partial \theta} \right\| \leq \int_{\underline{v}}^{v} \left\| \frac{\partial f(t, \theta)}{\partial \theta} \right\| dt \leq 2B^{\frac{1}{2}} \int_{\underline{v}}^{v} \|P(t)\| \sqrt{\phi(t)} dt \leq 2B(v - \underline{v})\zeta_n$$

where the last inequality uses the assumption that $\phi(v)$ is bounded. Similarly, using the

ϕ-orthonormality of the basis functions, we obtain the additional bound

$$\begin{aligned}\left\|\frac{\partial F(v,\theta)}{\partial\theta}\right\| &\le \int_{\underline{v}}^{v}\left\|\frac{\partial f(t,\theta)}{\partial\theta}\right\| dt\\ &\le 2B^{\frac{1}{2}}\int_{\underline{v}}^{v}\|P(t)\|\sqrt{\phi(t)}dt\\ &\le 2B(v-\underline{v})^{\frac{1}{2}}\int_{\underline{v}}^{v}\|P(t)\|^2\phi(t)dt\\ &= 2B(v-\underline{v})^{\frac{1}{2}}\sqrt{K_L}.\end{aligned}$$

where the third inequality uses the assumption that $\phi(v)$ is bounded away from zero and Jensen's inequality[9] and the last equality uses the ϕ-orthonormality of the basis functions. Combining the last two inequalities shows that

$$\left\|\frac{\partial F(v,\theta)}{\partial\theta}\right\| \le 2B\min\left\{\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}, (v-\underline{v})\zeta_L\right\}. \tag{B.53}$$

For the final auxiliary inequality,

$$\left\|\frac{\partial^2 F(v,\theta)}{\partial\theta\partial\theta'}\right\| \le 2B\left(\left(\int_{\underline{v}}^{v}\|P(t)\|^2\phi(t)dt\right)^{\frac{1}{2}}\right)^2 \le 2B(v-\underline{v})^{\frac{1}{2}}\zeta_L\sqrt{K_L}. \tag{B.54}$$

We can now establish the inequalities of the lemma. The first inequality follows directly by equation (B.52) as

$$|h(v,\theta,p)| \le \frac{B^p(v-\underline{v})^p}{\frac{1}{B^{p-1}}\int_{\underline{v}}^{v}(t-\underline{v})^{p-1}dt} = pB^{2p-1}.$$

To establish the second inequality, note

$$\begin{aligned}h_\theta(v,\theta,p) &= \frac{pF(v,\theta)^{p-1}\frac{\partial F(v,\theta)}{\partial\theta}}{\int_{\underline{v}}^{v}F(t,\theta)^{p-1}dt} - h(v,\theta,p)\frac{\int_{\underline{v}}^{v}(p-1)F(t,\theta)^{p-2}\frac{\partial F(t,\theta)}{\partial\theta}}{\int_{\underline{v}}^{v}F(t,\theta)^{p-1}dt}\\ &\equiv \psi_1(v,\theta,p) - h(v,\theta,p)\psi_2(v,\theta,p).\end{aligned}$$

---

[9] To see this, note $|f(x)| = \sqrt{f(x)^2}$ so that

$$\int_{\underline{v}}^{v}|f(x)|dx = (v-\underline{v})\int_{\underline{v}}^{v}\frac{\sqrt{f(x)^2}}{v-\underline{v}}dx \le (v-\underline{v})\left(\frac{1}{v-\underline{v}}\int_{\underline{v}}^{v}f(x)^2dx\right)^{\frac{1}{2}} \le (v-\underline{v})^{\frac{1}{2}}\left(\int_{\underline{v}}^{\overline{v}}f(x)^2dx\right)^{\frac{1}{2}}$$

where the first inequality uses Jensen's inequality (where the expectation taken with respect to the uniform distribution on $[\underline{v},v]$) and the second inequality follows as we are increasing the domain over which we are integrating a positive function.

Using equation (B.52) and the inequality $\left\|\frac{\partial F(v,\theta)}{\partial\theta}\right\| \leq 2B\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}$ derived from equation (B.53),

$$\|\psi_1(v,\theta,p)\| \leq 2pB^{2p-1}\sqrt{K_L}\frac{(v-\underline{v})^{p-\frac{1}{2}}}{\int_{\underline{v}}^{v}(t-\underline{v})^{p-1}dt} = 2p^2B^{2p-1}\sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}}$$

and

$$\|\psi_2(v,\theta,p)\| \leq 2(p-1)C^{2p-2}\sqrt{K_n}\frac{\int_{\underline{v}}^{v}(t-\underline{v})^{p-\frac{1}{2}}dt}{\int_{\underline{v}}^{v}(t-\underline{v})^{p-1}dt} = 2(p-1)B^{2p-2}\frac{p}{p-\frac{1}{2}}\sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}}$$

Hence, by the triangle inequality and the first inequality of the lemma

$$\|h_\theta(v,\theta,p)\| \leq 2p^2B^{2p-1}\left(1+\frac{(p-1)B^{2p-2}}{p-\frac{1}{2}}\right)\sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}}. \quad \text{(B.55)}$$

Using a similar argument with the bound $\left\|\frac{\partial F(v,\theta)}{\partial\theta}\right\| \leq 2B\zeta_L(v-\underline{v})$ yields the inequality

$$\|h_\theta(v,\theta,p)\| \leq 2p^2B^{2p-1}\left(1+(p-1)B^{2p-2}\right)\zeta_L. \quad \text{(B.56)}$$

As C can be chosen to be larger than the two multiplicative constants appearing in the bounds (B.55) and (B.56), these two equations establish the third inequality of the lemma. To establish the sixth inequality, notice

$$\psi_{1\theta}(v,\theta,p) = \frac{pF(v,\theta)^{p-1}}{\int_{\underline{v}}^{v}F(t,\theta)^{p-1}dt}\left(\frac{\partial^2F(v,\theta)}{\partial\theta\partial\theta'} + (p-1)\frac{1}{F(v,\theta)}\frac{\partial F(v,\theta)}{\partial\theta}\frac{\partial F(v,\theta)}{\partial\theta'}\right.$$
$$\left. -(p-1)\frac{\partial F(v,\theta)}{\partial\theta}\frac{\int_{\underline{v}}^{v}F(t,\theta)^{p-2}\frac{\partial F(t,\theta)}{\partial\theta}dt}{\int_{\underline{v}}^{v}F(t,\theta)^{p-1}dt}\right)$$

So the triangle inequality and the previous bounds give

$$\|\psi_{1\theta}(v,\theta,p)\| \leq 2p^2B^{2p-1}\left(1+4(p-1)C^2+(p-1)4B^{2p-2}\right)\sqrt{K_L}\zeta_L(v-\underline{v})^{-\frac{1}{2}}$$

where we have used equation (B.55) to bound the first term appearing in the parenthesis on the right-hand side and equation (B.54) to bound the second and third terms.[10] Using similar

[10]For example, to bound the second term in the parenthesis, we have

$$(p-1)\frac{1}{F(v,\theta)}\left\|\frac{\partial F(v,\theta)}{\partial\theta}\right\|\left\|\frac{\partial F(v,\theta)}{\partial\theta}\right\| \leq \frac{(p-1)C}{v-\underline{v}}\left(2B\zeta_L(v-\underline{v})\right)\left(2B\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}\right)$$
$$= 4(p-1)B^3\sqrt{K_L}\zeta_n(v-\underline{v})^{\frac{1}{2}}.$$

arguments, we can establish

$$\|\psi_{2\theta}(v,\theta,p)\| \leq 2B^{2p-2}\frac{p(p-1)}{p-\frac{1}{2}}\left(1+2(p-2)+2B^{2p-2}(p-1)\right)\zeta_L\sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}}.$$

Consequently, by the previous derivations, for a sufficiently large $C<\infty$

$$\begin{aligned}\|h_{\theta\theta}(v,\theta,p)\| &\leq \|\psi_{1\theta}(v,\theta,p)\| + |h(v,\theta,p)|\|\psi_{2\theta}(v,\theta,p)\| + \|\psi_2(v,\theta,p)\|\|h_\theta(v,\theta,p)\| \\ &\leq C\zeta_L\sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}}\end{aligned}$$

where we have used the first and third inequalities of the lemma combined with the observation that

$$\|\psi_2(v,\theta,p)\| \leq 2(p-1)B^{2p-2}\zeta_L$$

which establishes the sixth inequality. □

Lemma B.18. *Suppose Assumption (B.3) holds and let $\theta \in \Theta_{K_L}$ and $b \in (\underline{v}, \overline{b}(\theta,p))$ and . There exists a unique $v(b,\theta,p) \in (\underline{v},\overline{v})$ such that*

$$v(b,\theta,p) - \frac{\int_{\underline{v}}^{v(b,\theta,p)} F(t,\theta)^{p-1}dt}{F(v,\theta)^{p-1}} = b.$$

Moreover, there exists a constant $C<\infty$ not depending on b or θ such that

$$\begin{aligned}\left\|\frac{dv(b,\theta,p)}{d\theta}\right\| &\leq C\min\left\{\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}, \zeta_L(v-\underline{v})\right\}\Big|_{v=v(b,\theta,p)} \\ \left\|\frac{d^2v(b,\theta,p)}{d\theta d\theta'}\right\| &\leq C\min\left\{\sqrt{K_L}\zeta_L(v-\underline{v})^{\frac{1}{2}}, \zeta_L^2(v-\underline{v})\right\}\Big|_{v=v(b,\theta,p)} \\ \left\|\frac{d^3v(b,\theta,p)}{d\theta d\theta d\theta_i}\right\| &\leq C\zeta_L^2\end{aligned}$$

where the last inequality holds for all $i \in \{1,...,K_L\}$.

Proof. For the proof, let B denote the maximum of the constants in Assumption (B.3) and lemmas (B.17) and (B.19). Furthermore, let

$$s(v,b,\theta,p) \equiv v - b - \frac{\int_{\underline{v}}^{v} F(t,\theta)^{p-1}dt}{F(v,\theta)^{p-1}}.$$

It suffices to show $s(\cdot, b, \theta, p)$ has a unique zero. By L'Hospital's rule, the assumption that $f(\underline{v}, \theta) > \frac{1}{B}$ and the definition of $\overline{b}(\theta, p)$, $s(\cdot, b, \theta, p)$ is a continuous and strictly increasing function with range-space $[\underline{v}, \overline{b}(\theta, p)]$. Existence of a zero to $s(\cdot, b, \theta, p)$ follows from the choice of b indicated in the lemma and the intermediate value theorem. Uniqueness follows from strict monotonicity of $s(v, b, \theta, p)$.

For the second part of the lemma, note for any $b \in (\underline{v}, \overline{b}(\theta, p))$, we have $v(b, \theta, p) \in (\underline{v}, \overline{v})$ and $s_v(\cdot, b, \theta, p) > 0$ on $(\underline{v}, \overline{v})$. Therefore, the conditions of the implicit function theorem apply and we have

$$\frac{dv(b,\theta,p)}{d\theta} = \frac{F(v,\theta)}{f(v,\theta)} \frac{\int_{\underline{v}}^{v} F(t,\theta)^{p-2} \frac{\partial F(t,\theta)}{\partial \theta} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} - \frac{\frac{\partial F(v,\theta)}{\partial \theta}}{f(v,\theta)}.$$

Therefore by the triangle inequality and the bounds of lemmas the (B.17) and (B.19),

$$\left\| \frac{dv(b,\theta,p)}{d\theta} \right\| \leq B(B+1) \min \left\{ \sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}, \zeta_L(v-\underline{v}) \right\}.$$

Letting $C > B(B+1)$ gives the desired result. □

Lemma B.19. *Under Assumptions (B.2) and (B.3), there exists* $0 < B < \infty$ *such that the following inequalities hold*

1. $F(v, \theta) \leq B(v - \underline{v})$
2. $F(v, \theta) \geq \frac{1}{B}(v - \underline{v})$
3. $\left\| \frac{dF(v,\theta)}{d\theta} \right\| \leq B \max \left\{ \zeta_L, \sqrt{K_L}(v-\underline{v})^{\frac{1}{2}} \right\}$
4. $\left\| \frac{d^2 F(v,\theta)}{d\theta d\theta'} \right\| \leq B \max \left\{ \zeta_L^2, \sqrt{K_L}\zeta_L(v-\underline{v})^{\frac{1}{2}} \right\}.$

Proof. Let C be the constant in Assumption (B.3). The first two inequalities follow as

$$F(v,\theta) = \int_{\underline{v}}^{v} f(s,\theta) ds = (v-\underline{v}) f(\tilde{v},\theta) \in \left] \frac{1}{C}(v-\underline{v}) \, , \, C(v-\underline{v}) \right]$$

where the second equality uses the mean value theorem. For the third inequality, notice

$$\frac{dF(v,\theta)}{d\theta} = \int_{\underline{v}}^{v} \frac{df(t,\theta)}{d\theta}dt = 2\int_{\underline{v}}^{v}(P(t)'\theta)P(t)\phi(t)dt.$$

Therefore, as $|P(v)'\theta| = \sqrt{\frac{f(v,\theta)}{\phi(v)}} < C$ we have $\left\|\frac{dF(v,\theta)}{d\theta}\right\| \leq 2C\int_{\underline{v}}^{v}\|P(t)\|\phi(t)dt$. Notice,

$$\begin{aligned}\int_{\underline{v}}^{v}\|P(t)\|\phi(t)dt &= \int_{\underline{v}}^{v}\sqrt{\|P(t)\|^2\phi(t)^2}dt \\ &\leq (v-\underline{v})^{\frac{1}{2}}\sqrt{\int_{\underline{v}}^{v}\|P(t)\|^2\phi(t)^2dt} \\ &\leq C\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}\end{aligned}$$

and

$$\int_{\underline{v}}^{v}\|P(t)\|\phi(t)dt \leq \zeta_L\int_{\underline{v}}^{v}\phi(t)dt \leq C\zeta_L(v-\underline{v}).$$

Combining the last three equations yields

$$\left\|\frac{dF(v,\theta)}{d\theta}\right\| \leq 2C^2\min\left\{\zeta_L(v-\underline{v})\,,\ \sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}\right\}.$$

Finally, as $\frac{d^2f(v,\theta)}{d\theta d\theta'} = 2P(v)P(v)'\phi(v)$, we have

$$\left\|\frac{d^2F(v,\theta)}{d\theta d\theta'}\right\| \leq 2\int_{\underline{v}}^{v}\|P(t)\|^2\phi(t)dt,$$

Therefore, trivially by the definition of ζ_L, $\left\|\frac{d^2F(v,\theta)}{d\theta d\theta'}\right\| \leq C\zeta_L^2(v-\underline{v})$. By previous arguments we also have

$$\int_{\underline{v}}^{v}\|P(t)\|^2\phi(t)dt \leq \zeta_L\int_{\underline{v}}^{v}\|P(t)\|\phi(t)dt \leq C\zeta_L\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}}.$$

Hence,

$$\left\|\frac{d^2F(v,\theta)}{d\theta d\theta'}\right\| \leq 2C^2\min\left\{\zeta_L^2(v-\underline{v})\,,\ \zeta_L\sqrt{K_L}(v-\underline{v})\right\}.$$

So letting,

$$B = \max\left\{C\,,\ 2C^2\right\}$$

establishes the lemma. □

Lemma B.20. *Suppose assumption (B.3) holds. Then for any* $\theta \in \Theta$*,* $v \in [\underline{v}, \overline{v}]$ *and* $p \in \{2, ..., \overline{p}\}$ *and let*

$$s(v, \theta, b, p) = (v - b) - \frac{\int_{\underline{v}}^{v} F(t, \theta)^{p-1} dt}{F(v, \theta)^{p-1}}.$$

Then there exists a constant $0 < B < \infty$ *with* $\frac{1}{B} < s'(v, \theta, b, p) < B$.

Proof. Let C be the constant defined in Assumption (B.3) and note

$$s'(v, \theta, b, p) = \frac{(p-1) f(v, \theta) \int_{\underline{v}}^{v} F(t, \theta)^{p-1} dt}{F(v, \theta)^{p}}.$$

Therefore, as $\frac{1}{C}(t - \underline{v}) < F(t, \theta) < C(t - \underline{v})$ and we have

$$s'(v, \theta, b, p) \leq (p-1) C^{2p} \frac{\int \underline{v}^{v} (t - \underline{v})^{p-1} dt}{(v - \underline{v})^{p}} = \frac{p-1}{p} C^{2p} \leq 2C^{2\overline{p}}.$$

Similarly,

$$s'(v, \theta, b, p) \geq (p-1) \frac{1}{C^{2p}} \frac{\int_{\underline{v}}^{v} (t - \underline{v})^{p-1} dt}{(v - \underline{v})^{p}} = \frac{p-1}{p} \frac{1}{C^{2p}} \geq C^{-2\overline{p}}.$$

Therefore, letting $B = 2C^{2\overline{p}}$ gives the desired result. □

Lemma B.21. *For any* $v \in [\underline{v}, \overline{v}]$ *and* $p \in \{2, ..., \overline{p}\}$*, let*

$$\psi_1(v, \theta, p) = \frac{\int_{\underline{v}}^{v} \frac{dF(t,\theta)}{d\theta} F(t, \theta)^{p-2} dt}{\int_{\underline{v}}^{v} F(t, \theta)^{p-1} dt}$$

$$\psi_2(v, \theta, p) = \psi_1(v, \theta, p) \frac{-F(v, \theta)^{p}}{f(v, \theta) \int_{\underline{v}}^{v} F(t, \theta)^{p-1} dt}$$

$$\psi_3(v, \theta, p) = \frac{F(v, \theta)^{p-1} \frac{dF(v,\theta)}{d\theta}}{\int_{\underline{v}^{v}} F(t, \theta)^{p-1} dt}$$

Under Assumption (B.3), there exists $0 < B < \infty$ *such that the following inequalities hold*

1. $\|\psi_j(v, \theta, p)\| \leq B \min\left\{\zeta_L \, , \, \sqrt{K_L}(v - \underline{v})^{-\frac{1}{2}}\right\}$
2. $\left\|\frac{\partial \psi_j(v,\theta,p)}{\partial \theta}\right\| \leq B \min\left\{\zeta_L^2 \, , \, \zeta_L \sqrt{K_L}(v - \underline{v})^{-\frac{1}{2}}\right\}$

3. $\left\| \frac{\partial^2 \psi_j(v,\theta,p)}{\partial\theta\partial\theta_i} \right\| \leq B\zeta_L^2(v-\underline{v})^{-\frac{1}{2}}$

for all $j \in \{1,2,3\}$ *where the last inequality holds for all* $i \in \{1,...,K_L\}$.

Proof. As the remaining inequalities follow from similar arguments, we only establish the inequalities for $j=1$. Let C be the maximum of the constants defined in Assumption (B.3) and lemma (B.19). By the bounds in lemma (B.19) we have

$$\begin{aligned}
\|\psi_1(v,\theta,p)\| &\leq \frac{\int_{\underline{v}}^{v} \left\| \frac{dF(t,\theta)}{d\theta} \right\| F(t,\theta)^{p-2} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} \\
&\leq \frac{C^{2p-2} \int_{\underline{v}}^{v} (t-\underline{v})^{p-2} \min\left\{ \zeta_L(t-\underline{v}) \, , \, \sqrt{K_L}(t-\underline{v})^{\frac{1}{2}} \right\} dt}{\int_{\underline{v}}^{v} (t-\underline{v})^{p-1} dt} \\
&\leq pC^{2p-2} \min\left\{ \zeta_L \, , \, \sqrt{K_L}(v-\underline{v})^{-\frac{1}{2}} \right\}
\end{aligned}$$

where the final inequality evaluates the integral on the denominator and uses the fact that $(t-\underline{v}) \leq (v-\underline{v})$ for all t in the domain of integration to bound the numerator. To establish the second bound, notice

$$\begin{aligned}
\frac{\partial\psi_1(v,\theta,p)}{\partial\theta'} &= \frac{\int_{\underline{v}}^{v} \frac{d^2F(t,\theta)}{d\theta d\theta'} F(t,\theta)^{p-2} + (p-2)\frac{dF(t,\theta)}{d\theta}\frac{dF(t,\theta)}{d\theta'} F(t,\theta)^{p-3} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} \\
&\quad - \frac{\int_{\underline{v}}^{v} \frac{dF(t,\theta)}{d\theta} F(t,\theta)^{p-2} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} \frac{\int_{\underline{v}}^{v} \frac{dF(t,\theta)}{d\theta} F(t,\theta)^{p-2} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt}. \qquad \text{(B.57)}
\end{aligned}$$

To bound $\left\| \frac{\partial\psi_1(v,\theta,p)}{\partial\theta'} \right\|$, note by the bounds in lemma (B.19) for any $t \in [\underline{v}, v]$, we have

$$\begin{aligned}
&\left\| \frac{d^2F(t,\theta)}{d\theta d\theta'} F(t,\theta)^{p-2} + (p-2)\frac{dF(t,\theta)}{d\theta}\frac{dF(t,\theta)}{d\theta'} F(t,\theta)^{p-3} \right\| \\
&\quad \leq C^{p-2} \Big((t-\underline{v})^{p-2} \min\left\{ \zeta_L^2(t-\underline{v}) \, , \, \zeta_L\sqrt{K_L}(t-\underline{v})^{\frac{1}{2}} \right\} \\
&\qquad + (p-2)\zeta_L(t-\underline{v})^{p-2} \min\left\{ \zeta_L(t-\underline{v}) \, , \, \sqrt{K_L}(t-\underline{v})^{\frac{1}{2}} \right\} \Big) \\
&\quad \leq (p-1)C^{p-2}(v-\underline{v})^{p-1} \min\left\{ \zeta_L^2 \, , \, \zeta_L\sqrt{K_L}(v-\underline{v})^{\frac{1}{2}} \right\}
\end{aligned}$$

and

$$\int_{\underline{v}}^{v} \left\| \frac{dF(t,\theta)}{d\theta} \right\| F(t,\theta)^{p-2} dt \leq C^{p-1} \int_{\underline{v}}^{v} (t-\underline{v})^{p-2} \min \left\{ \zeta_L (t-\underline{v}) \, , \, \sqrt{K_L} (t-\underline{v})^{\frac{1}{2}} \right\} dt$$
$$\leq C^{p-1} (v-\underline{v})^{p} \min \left\{ \zeta_L \, , \, \sqrt{K_L} (v-\underline{v})^{-\frac{1}{2}} \right\}$$

where the last inequality uses the fact that $(t-\underline{v}) \leq (v-\underline{v})$ for all t in the range of integration. Using the triangle and CauchySchwarz inequalities, equation (B.57), the previous two equations and some algebra yields,

$$\left\| \frac{\partial \psi_1(v,\theta,p)}{\partial \theta'} \right\| \leq \left(p(p-1) C^{2p-1} (v-\underline{v})^{p-1} p^2 C^{4p-4} \right) \min \left\{ \zeta_L^2 \, , \, \zeta_L \sqrt{K_L} (v-\underline{v})^{-\frac{1}{2}} \right\}.$$

Finally, notice

$$\begin{aligned}
&\frac{\partial^2 \psi_1(v,\theta,p)}{\partial \theta \partial \theta_i} \\
&\quad = \frac{(p-2) \int_{\underline{v}}^{v} \frac{d^2 F(t,\theta)}{d\theta d\theta'} \frac{dF(t,\theta)}{d\theta_i} F(t,\theta)^{p-3} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1}} + \frac{(p-2) \int_{\underline{v}}^{v} \frac{dF(t,\theta)}{d\theta} \frac{d^2 F(t,\theta)}{d\theta d\theta_i} + \frac{d^2 F(t,\theta)}{d\theta d\theta_i} \frac{dF(t,\theta)}{d\theta} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} \\
&\qquad + \frac{(p-2)(p-3) \int_{\underline{v}}^{v} \frac{dF(t,\theta)}{d\theta} \frac{dF(t,\theta)}{d\theta} \frac{dF(t,\theta)}{d\theta_i} F(t,\theta)^{p-4} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} \\
&\qquad - \frac{d\psi_1(v,\theta,p)}{d\theta} \left((p-1) \frac{\int_{\underline{v}}^{v} F(t,\theta)^{p-2} \frac{dF(t,\theta)}{d\theta_i} dt}{\int_{\underline{v}}^{v} F(t,\theta)^{p-1} dt} \right) \\
&\qquad - \psi_1(v,\theta,p) \left(\frac{d\psi_1(v,\theta,p)}{d\theta_i} \right)' - \left(\frac{d\psi_1(v,\theta,p)}{d\theta_i} \right) \psi_1(v,\theta,p)'.
\end{aligned}$$

For brevity, we only provide the bound for the first term as the remaining terms follow similar arguments. Notice,

$$\left\| \frac{dF(t,\theta)}{d\theta} \right\| \leq \int_{\underline{v}}^{t} 2 |P_i(s)(P(s)'\theta)\phi(s) ds| \leq 2C^2 \int_{\underline{v}}^{t} |P_i(s)| \sqrt{\phi(s)} ds \leq 2C^2 (t-\underline{v})^{\frac{1}{2}}$$

where the last inequality follows as

$$\begin{aligned}\int_{\underline{v}}^{t} |P_s(t)|\sqrt{\phi(s)}ds &= (t-\underline{v})\left(\frac{1}{t-\underline{v}}\int_{\underline{v}}^{v}\sqrt{P_i(s)^2\phi(s)}ds\right)\\ &\leq (t-\underline{v})\left(\frac{1}{t-\underline{v}}\int_{\underline{v}}^{t} P_i(s)^2\phi(s)\right)^{\frac{1}{2}}\\ &\leq (t-\underline{v})^{\frac{1}{2}}\end{aligned}$$

where the last inequality increases the upper bound of integration to $\overline{v}$ and uses the assumed ϕ-orthonormality of $P(\cdot)$. Using this inequality combined with the fact that $\left\|\frac{d^2F(t,\theta)}{d\theta d\theta'}\right\| \leq C(t-\underline{v})\zeta_L^2$ gives

$$\left\|\frac{\int_{\underline{v}}^{v}\frac{d^2F(t,\theta)}{d\theta d\theta'}\frac{dF(t,\theta)}{d\theta_i}F(t,\theta)^{p-3}dt}{\int_{\underline{v}}^{v}F(t,\theta)^{p-1}}\right\|$$

$$\leq 2C^{2p-1}\zeta_L^2\frac{\int_{\underline{v}}^{v}(t-\underline{v})^{p-\frac{3}{2}}dt}{\int_{\underline{v}}^{v}(t-\underline{v})^{p-1}dt} = 2C^{2p-1}\zeta_L^2\frac{p}{p-\frac{1}{2}} \leq 4C^{2p-1}\zeta_L^2\frac{p}{p-\frac{1}{2}}.$$

Deriving similar bounds for the remaining terms in the definition of $\frac{\partial^2\psi(v,\theta,p)}{\partial\theta\partial\theta_i}$ and using the triangle inequality establishes the bound.

□

Lemma B.22. *Let* $m_L(\theta) = \frac{1}{L}\sum_{i=1}^{L}\sum_{i=1}^{p_l}\left(\frac{d\log(g(b_{il},\theta,p_l))}{d\theta} - \mu(\theta,p_l)\right)$. *Under Assumptions (B.1) - (B.3), there exists a constant* $0 < B < \infty$ *such that*

$$\|m_L(\theta)\| \leq B\min\left\{\zeta_L\,,\ \sqrt{K_L}\left(1+\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\frac{1}{(v(b_{il},\theta,p_l)-\underline{v})^{\frac{1}{2}}}\right)\right\}$$

$$\left\|\frac{dm_L(\theta)}{d\theta'}\right\| \leq B\min\left\{\zeta_L^2\,,\ \sqrt{K_L}\zeta_L\left(1+\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\frac{1}{(v(b_{il},\theta,p_l)-\underline{v})^{\frac{1}{2}}}\right)\right\}$$

$$\left\|\frac{d^2m_L(\theta)}{d\theta' d\theta_i}\right\| \leq B\sqrt{K_L}\zeta_L^2\left(1+\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\frac{1}{(v(b_{il},\theta,p_l)-\underline{v})^{\frac{1}{2}}}\right)$$

where $v(b,\theta,p)$ *denotes the unique value such that* $0 = (v(b,\theta,p)-b) - \frac{\int_{\underline{v}}^{v(b,\theta,p)}F(t,\theta)^{p-1}dt}{F(v(b,\theta,p),\theta)^{p-1}}$.

Proof. The proof of these inequalities follows by appealing to several previously established lemmas (B.16), (B.18), (B.21) and (B.23). Let C be the maximum of the constants appearing in those lemmas and the constant in Assumption (B.3). For brevity, we only show how to combine the results of previous lemmas to prove the two statements. Let C be a constant larger than the constants appearing in lemmas , and note

$$\begin{aligned}\|m_L(\theta)\| &\leq \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\left\|\frac{d\log(g(v(b_{il},\theta,p_l),\theta,p_l)}{d\theta}\right\| + \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p}\|\mu(\theta,p_l)\| \\ &\leq C^2\frac{1}{L}\sum_{l}^{L}\sum_{i=1}^{p_l}\zeta_L + \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}C\zeta_L \\ &\leq \overline{p}C(C+1)\zeta_L\end{aligned}$$

where we have used lemma (B.16) to get

$$\frac{d\log(g(b,\theta,p))}{d\theta} = \frac{\frac{dg(b,\theta,p)}{d\theta}}{g(b,\theta,p)} \leq C^2\zeta_L.$$

Similarly,

$$\begin{aligned}\|m_L(\theta)\| &\leq \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}\left\|\frac{d\log(g(v(b_{il},\theta,p_l),\theta,p_l)}{d\theta}\right\| + \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p}\|\mu(\theta,p_l)\| \\ &\leq C^2\frac{1}{L}\sum_{l}^{L}\sum_{i=1}^{p_l}\sqrt{K_L}(v(b_{il},\theta,p_l)-\underline{v})^{-\frac{1}{2}} + \frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p_l}C\sqrt{K_L}(\overline{v}-\underline{v})^{\frac{1}{2}} \\ &\leq C^2\sqrt{K_L}\left(\overline{p}+\frac{1}{L}\sum_{l=1}^{l}\sum_{i=1}^{p_l}(v(b_{il},\theta,p)-\underline{v})^{-\frac{1}{2}}\right)\end{aligned}$$

where we have used lemma (B.16) to get

$$\frac{d\log(g(b,\theta,p))}{d\theta} = \frac{\frac{dg(b,\theta,p)}{d\theta}}{g(b,\theta,p)} \leq C^2\sqrt{K_L}(v(b,\theta,p)-\underline{v})^{-\frac{1}{2}}.$$

This establishes the first inequality in the lemma. □

Lemma B.23. *For any* $p \in \{2,...,\overline{p}\}$*, let* $\mu(\theta,p) = -\frac{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-2}\frac{dF(t,\theta)}{d\theta}dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt}$. *Under Assumption (B.3), there exists* $B < \infty$ *such that*

$$\begin{aligned}\|\mu(\theta,p)\| &\leq B\min\{\zeta_L\,,\,\sqrt{K_L}\} \\ \left\|\frac{d\mu(\theta,p)}{d\theta'}\right\| &\leq B\min\{\zeta_L^2\,,\,\zeta_L\sqrt{K_L}\} \\ \left\|\frac{d^2\mu(\theta,p)}{d\theta' d\theta_i}\right\| &\leq B\zeta_L^2\end{aligned}$$

where the last inequality holds for all $i \in \{1, ..., K_L\}$*.*

Proof. Let C be a constant larger than the constants appearing in Assumption (B.3) and lemma (B.19). Then,

$$\begin{aligned}
\|\mu(\theta, p)\| &\leq \frac{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-2} \left\| \frac{dF(t,\theta)}{d\theta} \right\| dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1} dt} \\
&\leq pC^{2p-1} \frac{\int_{\underline{v}}^{\overline{v}} (t-\underline{v})^{p-2} \left\| \frac{dF(t,\theta)}{d\theta} \right\| dt}{(v-\underline{v})^p} \\
&\leq pC^{2p+2} \min\left\{ \zeta_L, \sqrt{K_L}(\overline{v}-\underline{v})^{-\frac{1}{2}} \right\}. \qquad \text{(B.58)}
\end{aligned}$$

where the last inequality uses

$$\begin{aligned}
\left\| \frac{dF(t,\theta)}{d\theta} \right\| &\leq \int_{\underline{v}}^{t} |P(s)'\theta| \|P(s)\| \phi(s) ds \\
&\leq C \int_{\underline{v}}^{t} \|P(s)\| \phi(s) ds \\
&\leq C^2 \min\{\zeta_L, \sqrt{K_L}(t-\underline{v})^{\frac{1}{2}}\}
\end{aligned}$$

which uses the definition of ζ_L and the fact that by Hölder's inequality

$$\begin{aligned}
\int_{\underline{v}}^{t} \|P(s)\| \phi(s) ds &\leq C \int_{\underline{v}}^{t} \|P(s)\| \sqrt{\phi(s)} ds \\
&\leq C(t-\underline{v})^{\frac{1}{2}} \left(\int_{\underline{v}}^{\overline{v}} \|P(s)\|^2 \phi(s) ds \right)^{\frac{1}{2}} \\
&= C(t-\underline{v})^{\frac{1}{2}} \sqrt{K_L}.
\end{aligned}$$

Further, notice

$$\begin{aligned}
\frac{d\mu(\theta,p)}{d\theta'} =& \frac{\int_{\underline{v}}^{\overline{v}} (p-2) F(t,\theta)^{p-3} \frac{dF(t,\theta)}{d\theta} \frac{dF(t,\theta)}{d\theta} + F(t,\theta)^{p-2} \frac{d^2 F(t,\theta)}{d\theta d\theta} dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1} dt} \\
&- \mu(\theta,p) \frac{(p-1) \int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-2} \frac{dF(t,\theta)}{d\theta'} dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1} dt}
\end{aligned}$$

We then proceed to use the triangle inequality and bound both terms above. As the remaining bounds follow similar arguments, we only establish the first bound. Notice, using similar techniques as in establishing the bounds in equation, we can find a constant C_1 such that (B.58),

$$\left\| \frac{\int_{\underline{v}}^{\overline{v}} (p-2)F(t,\theta)^{p-3} \frac{dF(t,\theta)}{d\theta}\frac{dF(t,\theta)}{d\theta'} + F(t,\theta)^{p-2}\frac{d^2F(t,\theta)}{d\theta d\theta} dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt} \right\| \leq C_1 \min\left\{ \zeta_L^2 \, , \, \sqrt{K_L}\zeta_L \frac{1}{(\overline{v}-\underline{v})^{\frac{1}{2}}} \right\}.$$

Therefore, we can find a constant C_2 such that

$$\left\| \frac{d\mu(\theta,p)}{d\theta} \right\| \leq C_2 \min\left\{ \zeta_L^2 \, , \, \zeta_L \sqrt{K_L} \frac{1}{(\overline{v}-\underline{v})^{\frac{1}{2}}} \right\}. \tag{B.59}$$

To establish the final bound notice,

$$\begin{aligned}
\frac{d^2\mu(\theta,p)}{d\theta' d\theta_i} &= \frac{\int_{\underline{v}}^{\overline{v}} (p-2)F(t,\theta)^{p-3} \left(\frac{dF(t,\theta)}{d\theta}\frac{d^2F(t,\theta)}{d\theta' d\theta_i} + \frac{d^2F(t,\theta)}{d\theta d\theta_i}\frac{dF(t,\theta)}{d\theta} \right)}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt} \\
&+ \frac{\int_{\underline{v}}^{\overline{v}} (p-3)\frac{dF(t,\theta)}{d\theta}\frac{dF(t,\theta)}{d\theta'}\frac{dF(t,\theta)}{d\theta_i} F(t,\theta)^{p-4}dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt} \\
&- \frac{\mu(\theta,p)(p-1)\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-2}\frac{d^2F(t,\theta)}{d\theta' d\theta_i} + (p-2)F(t,\theta)^{p-3}\frac{dF(t,\theta)}{d\theta'}\frac{dF(t,\theta)}{d\theta_i}dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt} \\
&- \frac{d\mu(\theta,p)}{d\theta_i}(p-1)\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-2}\frac{dF(t,\theta)}{d\theta'}dt \int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt \\
&- \frac{d\mu(\theta,p)}{d\theta'}\frac{(p-1)\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-2}\frac{dF(t,\theta)}{d\theta_i}dt}{\int_{\underline{v}}^{\overline{v}} F(t,\theta)^{p-1}dt}.
\end{aligned}$$

As the remaining bounds follow similar arguments, we only bound the first term,

$$\begin{aligned}
&\left\|\frac{d^2\mu(\theta,p)}{d\theta' d\theta_i}\right\| \\
&\leq pC^{p-1}\frac{\int_{\underline{v}}^{\overline{v}}(p-2)2C^{p-2}\zeta_L\left\|\frac{d^2F(t,\theta)}{d\theta d\theta_i}\right\|(t-\underline{v})^{p-2}+(p-3)C^{p-2}\zeta_L^2(t-\underline{v})^{p-2}\left|\frac{dF(t,\theta)}{d\theta_i}\right|dt}{(v-\underline{v})^p} \\
&\leq 2pC^{2p-1}\frac{\int_{\underline{v}}^{\overline{v}}(p-2)4C^{2p-1}\zeta_L^2\sqrt{K_L}(t-\underline{v})^{p-\frac{3}{2}}+2(p-3)C^p\zeta_L^2(t-\underline{v})^{p-\frac{3}{2}}dt}{(v-\underline{v})^p} \\
&= 2pC^{2p-1}\zeta_L^2\sqrt{K_L}\left((p-2)4C^{2p-1}+2(p-3)C^p\right)\frac{\int_{\underline{v}}^{\overline{v}}(t-\underline{v})^{p-\frac{3}{2}}dt}{(v-\underline{v})^p}\zeta_L^2\sqrt{K_L} \\
&= 2pC^{2p-1}\left((p-2)4C^{2p-1}+2(p-3)C^p\right)\frac{1}{p-\frac{1}{2}}\frac{1}{(\overline{v}-\underline{v})^{\frac{1}{2}}}\zeta_L^2\sqrt{K_L} \\
&\leq C_3\zeta_L^2\sqrt{K_L} \qquad \text{(B.60)}
\end{aligned}$$

for some constant $C_1 < \infty$ where we have used

$$\begin{aligned}
\left\|\frac{d^2F(t,\theta)}{d\theta d\theta_i}\right\| &\leq \int_{\underline{v}}^{t} 2|P_i(s)|\|P(s)\|\phi(s)ds \\
&\leq 2C\zeta_L\int_{\underline{v}}^{t}|P_i(s)|\sqrt{\phi(s)}ds \\
&\leq 2C\zeta_L\sqrt{K_L}(t-\underline{v})^{\frac{1}{2}}
\end{aligned}$$

where the last inequality uses the definition of ζ_L and an application of Hölder's inequality as used above. Additionally, we used the following inequality in expression (B.60)

$$\left|\frac{dF(t,\theta)}{d\theta_i}\right| \leq 2C^2\int_{\underline{v}}^{t}|P_i(s)|\sqrt{\phi(s)}ds \leq 2C^2(t-\underline{v})^{\frac{1}{2}}\sqrt{K_L}$$

where we have, again, used Hölder's inequality and the ϕ-orthonormality o $P(v)$. Bounding the remaining terms in a similar manner, we can use the triangle inequality to find a constant $C_4 < \infty$ such that

$$\left\|\frac{d^2\mu(\theta,p)}{d\theta' d\theta_i}\right\| \leq C_4\zeta_L^2\sqrt{K_L}. \qquad \text{(B.61)}$$

Letting B be larger than the constants appearing in equations (B.58), (B.59) and (B.61) gives the desired result. □

Lemma B.24. *Let $\theta_1, \theta_2 \in \Theta$ and suppose Assumption (B.3) holds. For any $p \in \{2, ..., \overline{p}\}$ and any $b \in [\underline{v}, \overline{b}_0(p)]$ define v_i to be the unique value with*

$$0 = s(v_i, \theta_i, b, p) \equiv (v_i - b) - \frac{\int_{\underline{v}}^{v_i} F(t, \theta_i)^{p-1} dt}{F(v_i, \theta_i)^{p-1}}$$

for $b \in [\underline{v}, \overline{b}(\theta_i, p)]$ and $v_i = \overline{v}$ otherwise for $i = \{1, 2\}$. There exists $0 < B < \infty$ such that

$$|v_1 - v_2| \leq B\zeta_L \|\theta_1 - \theta_2\|.$$

Proof. Let C denote the maximum of the constants appearing in Assumption (B.3) and lemmas (B.19) and (B.21). Without loss of generality, assume $\overline{b}(\theta_1, p) \leq \overline{b}(\theta_2, p)$. Suppose $b \in [\underline{v}, \overline{b}(\theta_1, p)]$. Notice,

$$\begin{aligned} s(v_1, \theta_2, b, p) &= (v_1 - b) - \frac{\int_{\underline{v}}^{v_1} F(t, \theta_1)^{p-1} dt}{F(v_1, \theta_1)^{p-1}} + \left(\frac{\int_{\underline{v}}^{v_1} F(t, \theta_2)^{p-1} dt}{F(v_1, \theta_2)^{p-1}} - \frac{\int_{\underline{v}}^{v_1} F(t, \theta_1)^{p-1} dt}{F(v_1, \theta_1)^{p-1}} \right) \\ &= \frac{\int_{\underline{v}}^{v_1} F(t, \theta_2)^{p-1} dt}{F(v_1, \theta_2)^{p-1}} - \frac{\int_{\underline{v}}^{v_1} F(t, \theta_1)^{p-1} dt}{F(v_1, \theta_1)^{p-1}} \\ &= \int_{\underline{v}}^{v_1} F(t, \theta_2)^{p-1} dt \left(\frac{1}{F(v_1, \theta_2)^{p-1}} - \frac{1}{F(v_1, \theta_1)^{p-1}} \right) \\ &\quad + \frac{1}{F(v_1, \theta_1)^{p-1}} \left(\int_{\underline{v}}^{v_1} F(t, \theta_2)^{p-1} - F(t, \theta_1)^{p-1} dt \right) \end{aligned}$$

where the first second equality follows by definition of $s(v_1, \theta_1, b, p) = 0$. Therefore, by the triangle inequality and the inequalities in lemmas (B.19) and (B.21), we have

$$|s(v_1, \theta_2, b, p)| \leq 2C^2 \zeta_L \|\theta_1 - \theta_2\| (v - \underline{v})$$

Suppose $s(v_1, \theta_2, b, p) \geq 0$ (the negative case is handled symmetrically, and thus omitted). As $s(\cdot, \theta, b, p)$ is an increasing function, this implies $v_1 > v_2$ Then,

$$\begin{aligned} 2C^2 \zeta_L \|\theta_1 - \theta_2\| (v - \underline{v}) &\geq s(v_1, \theta_2, b, p) - s(v_2, \theta_2, b, p) \\ &= \int_{v_2}^{v_1} s'(t, \theta_2, b, p) dt \\ &\geq (v_1 - v_2) \frac{1}{C} \end{aligned} \tag{B.62}$$

where the last inequality uses the fact that

$$s'(v,\theta,b,p) = \frac{(p-1)\int_{\underline{v}}^{v} F(t,\theta)^{p-1}dt}{F(v,\theta)^p f(v,\theta)} > \frac{1}{C}.$$

Re-arranging inequality (B.62) gives,

$$0 \leq v_1 - v_2 \leq 2C^3\zeta_L\|\theta_1-\theta_2\|.$$

Now suppose $b \in [\overline{b}(\theta_1,p), \overline{b}(\theta_2,p)]$. By definition, and the inequalities in lemma (B.24)

$$\begin{aligned}|\overline{b}(\theta_1,p)-\overline{b}(\theta_2,p)| &\leq \int_{\underline{v}}^{\overline{v}} |F(t,\theta_1)^{p-1} - F(t,\theta_2)^{p-1}|dt \\ &\leq C\zeta_L\|\theta_1-\theta_2\|.\end{aligned}$$

Letting $\eta(b,\theta,p)$ denote the function with $0 = s(\eta(b,\theta,p),\theta,p)$ the inverse-function theorem shows

$$\eta'(b,\theta,p) = \frac{1}{s'(v,\theta,p)} \leq C$$

where the last bound is due to lemma (B.20). Let $\tilde{v} = \eta(\overline{b}(\theta_1,p),\theta_2,p)$, so that $\tilde{v} \leq v_2$ and notice

$$\overline{v} - \tilde{v} \leq \int_{\overline{b}(\theta_1,p)}^{\overline{b}(\theta_2,p)} \eta'(t,\theta,p)dt \leq C|\overline{b}(\theta_2,p)-\overline{b}(\theta_1,p)| \leq C^2\zeta_L\|\theta_1-\theta_2\|.$$

As $v_1 = \overline{v}$ and $\tilde{v} \leq v_2$ the previous inequality yields

$$0 \leq v_1 - v_2 \leq \overline{v} - \tilde{v} \leq C^2\zeta_L\|\theta_1-\theta_2\|.$$

In the final case, if $b \geq \overline{b}(\theta_2,p)$ then $v_1 = v_2 = \overline{v}$. In any case, the desired inequality holds for suitably chosen C.

□

Lemma B.25. *Let*

$$\hat{V}_m \equiv \frac{1}{L}\sum_{l=1}^{L} m(\boldsymbol{b}_l,\hat{\theta}_{mm})m(\boldsymbol{b}_l,\hat{\theta}_{mm})'$$

Under Assumptions (B.1)-(B.5) we have $\|\hat{V}_m - Var(m(\boldsymbol{b},\theta_0))\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)$

Proof. First notice,

$$\begin{aligned}
\|m(\boldsymbol{b},\theta)m(\boldsymbol{b},\theta)'\|_F^2 &= \sum_{i=1}^{K_L}\sum_{j=1}^{K_L}|m_i(\boldsymbol{b},\theta)m_j(\boldsymbol{b},\theta)|^2 \\
&= \sum_{i=1}^{K_L}|m_i(\boldsymbol{b},\theta)|^2\sum_{j=1}^{K_L}|m_j(\boldsymbol{b},\theta)|^2 \\
&= \|m(\boldsymbol{b},\theta)\|^2
\end{aligned}$$

By Markov's inequality

$$P(\|\overline{X}-\mu\|_F > a) \leq \frac{\mathbb{E}_0\left(\|\overline{X}-\mu\|_F\right)}{a} \leq \frac{\sqrt{\mathbb{E}_0\left(\|\overline{X}-\mu\|^2\right)}}{a} \leq \frac{1}{\sqrt{L}}\frac{\sqrt{\mathbb{E}_0(\|X\|_F^2)}}{a}$$

where we have used the fact that for any random matrix A we have

$$\mathbb{E}_0\left(\|A-\mathbb{E}_0(A)\|_F^2\right) = \sum_{i=1}^{K_L}\sum_{j=1}^{K_L}\mathbb{E}_0\left(|A_{ij}-\mathbb{E}_0(A_{ij})|^2\right) \leq \mathbb{E}_0\left(\|A\|_F^2\right)$$

where the last line follows as

$$\mathbb{E}_0((A_{ij}-\mathbb{E}_0(A_{ij}))^2) = \mathbb{E}_0(A_{ij}^2) - \mathbb{E}_0(A_{ij})^2 \leq \mathbb{E}_0(A_{ij}^2).$$

Combining the last two statements shows,

$$\begin{aligned}
P\left(\left\|\frac{1}{L}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_0)m(\boldsymbol{b}_l,\theta_0)' - \mathbb{E}_0\left(m(\boldsymbol{b},\theta_0)m(\boldsymbol{b},\theta_0)'\right)\right\| > a\right) &\leq \frac{1}{\sqrt{L}}\frac{\sqrt{\mathbb{E}_0\left(\|m(\boldsymbol{b},\theta)\|^2\right)}}{a} \\
&\leq C\frac{\zeta_L^2}{\sqrt{L}}
\end{aligned}$$

Therefore, by Assumption (B.4) we have

$$\left\|\frac{1}{L}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_0)m(\boldsymbol{b}_l,\theta_0)' - \mathbb{E}_0\left(m(\boldsymbol{b},\theta_0)m(\boldsymbol{b},\theta_0)'\right)\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right).$$

Furthermore, notice

$$Var\left(\frac{1}{\sqrt{L}}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_0)\right) = \mathbb{E}_0\left(\|m(\boldsymbol{b},\theta_0)\|^2\right) - \mathbb{E}_0\left(m(\boldsymbol{b},\theta_0)\right)\mathbb{E}_0\left(m(\boldsymbol{b},\theta_0)\right)'.$$

Re-arranging this equation and applying lemma (B.10), gives

$$\left\|Var\left(\frac{1}{\sqrt{L}}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_0)\right)-\mathbb{E}_0\left(\|m(\boldsymbol{b},\theta_0)\|^2\right)\right\| \leq \|\mathbb{E}_0\left(m(\boldsymbol{b},\theta_0)\right)\|^2$$
$$= O_p\left(K_L^{1-2\gamma}\right)$$
$$= o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right).$$

Combining the last several statements shows

$$\left\|\frac{1}{L}\sum_{l=1}^{l}m(\boldsymbol{b}_l,\theta_0)m(\boldsymbol{b}_l,\theta_0)-Var(m(\boldsymbol{b},\theta_0))\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right). \tag{B.63}$$

Notice, for any $p \in \{2,...,\overline{p}\}$ and any $\boldsymbol{b} \in \mathbb{R}^p$ we have

$$\begin{aligned}&\left\|m(\boldsymbol{b},\theta_0)m(\boldsymbol{b},\theta_0)-m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\|\\ &\quad\leq \left\|m(\boldsymbol{b},\theta_0)\right\|\left\|m(\boldsymbol{b},\theta_0)-m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\|+\left\|m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\|\left\|m(\boldsymbol{b},\theta_0)-m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\|.\end{aligned} \tag{B.64}$$

By lemma (B.19), lemma (B.22) and a mean-value expansion, we have

$$\left\|m(\boldsymbol{b},\theta_0)\right\|\left\|m(\boldsymbol{b},\theta_0)-m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\| \leq C^2\frac{\zeta_L^2\sqrt{K_L}}{\sqrt{L}}\left\|\sqrt{L}\,(\hat{\theta}_{\mathrm{mm}}-\theta_0)\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)$$

for some constant $0 < B < \infty$ where we have used $\|\sqrt{L}\,(\hat{\theta}_{\mathrm{mm}}-\theta_0)\| = O_p(\sqrt{K_L})$ and the rate conditions in Assumption (B.4). Similarly, we have

$$\left\|m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\|\left\|m(\boldsymbol{b},\theta_0)-m(\boldsymbol{b},\hat{\theta}_{\mathrm{mm}})\right\| \leq C^2\frac{\zeta_L^2\sqrt{K_L}}{\sqrt{L}}\left\|\sqrt{L}\,(\hat{\theta}_{\mathrm{mm}}-\theta_0)\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right)$$

Plugging these bounds into equation (B.64) and using the triangle inequality and the rates in Assumption (B.4) gives

$$\left\|\frac{1}{L}\sum_{l=1}^{L}m(\boldsymbol{b}_l,\theta_0)m(\boldsymbol{b}_l,\theta_0)'-m(\boldsymbol{b}_l,\hat{\theta}_{\mathrm{mm}})m(\boldsymbol{b}_l,\hat{\theta}_{\mathrm{mm}})'\right\| = o_p\left(\varepsilon_L K_L^{-\frac{1}{2}}\right). \tag{B.65}$$

Therefore, by the triangle inequality and equations (B.63) and (B.65) gives

$$\begin{aligned}
&\left\| \frac{1}{L} \sum_{l=1}^{L} m(\boldsymbol{b}_l, \hat{\theta}_{\mathrm{mm}}) m(\boldsymbol{b}_l, \hat{\theta}_{\mathrm{mm}})' - Var(m(\boldsymbol{b}, \theta_0)) \right\| \\
&\leq \left\| \frac{1}{L} \sum_{l=1}^{L} m(\boldsymbol{b}_l, \hat{\theta}_{\mathrm{mm}}) m(\boldsymbol{b}_l, \hat{\theta}_{\mathrm{mm}})' - \frac{1}{L} \sum_{l=1}^{L} m(\boldsymbol{b}_l, \theta_0) m(\boldsymbol{b}_l, \theta_0)' \right\| \\
&\quad + \left\| \frac{1}{L} \sum_{l=1}^{l} m(\boldsymbol{b}_l, \theta_0) m(\boldsymbol{b}_l, \theta_0) - Var(m(\boldsymbol{b}, \theta_0)) \right\| \\
&= o_p\left(\varepsilon_L K_L^{-\frac{1}{2}} \right)
\end{aligned}$$

which establishes the desired result. □

Bibliography

Aarts, L. et al. (2007). Estimating the upper support point in deconvolution. *Scandinavian Journal of Statistics 34*(3), 552–568.

Ait-Sahalia, Y. and J. Duarte (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics 116*(1-2), 9–47.

Andrews, D. W. K. (1999). Estimation when a parameter is on a boundary. *Econometrica 67*(6), 1341–1383.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica 69*(3), 683–734.

Andrews, D. W. K. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Andrews, D. W. K. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Armstrong, T. and M. Kolesár (2016). Simple and honest confidence intervals in nonparametric regression. Working paper.

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

Athey, S. and J. Levin (2001). Information and competition in u.s. forest service timber auctions. *Journal of Political Economy 109*(2), 375–417.

Baldwin, L. et al. (1997). Bidder collusion at forest service timber sales. *Journal of Political Economy 105*(4), 657–699.

Bellec, P. C. (2016). Adaptive confidence sets in shape restricted regression. Working paper.

Beresteanu, A. (2005). Nonparametric Analysis of Cost Complementarities in the Telecommunications Industry. *RAND Journal of Economics 36*(4), 870–889.

Birke, M. and H. Dette (2007). Estimating a convex function in nonparametric regression. *Scandinavian Journal of Statistics 34*(2), 384–404.

Blundell, R., J. L. Horowitz, and M. Parey (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics 3*(1), 29–51.

Blundell, R., J. L. Horowitz, and M. Parey (2017). Nonparametric estimation of a nonseparable demand function under the slutsky inequality restriction. *The Review of Economics and Statistics 99*(2), 291–304.

Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics 26*(4), 607–616.

Cai, T. T., M. G. Low, and Y. Xia (2013). Adaptive confidence intervals for regression functions under shape constraints. *The Annals of Statistics 41*(2), 722–750.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2017). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association, forthcoming 113*, 767–779.

Chatterjee, S., A. Guntuboyina, and B. Sen (2015). On risk bounds in isotonic and other shape restricted regression problems. Working paper.

Chen, X., V. Chernozhukov, I. Fernández-Val, S. Kostyshak, and Y. Luo (2018). Shape-enforcing operators for point and interval estimators. Working paper.

Chen, X. and T. M. Christensen (2018). Optimal sup-normrates and uniforminference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics 9*(1), 39–84.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika 96*(3), 559–575.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*(2), 667–737.

Chernozhukov, V., W. K. Newey, and A. Santos (2015). Constrained conditional moment restriction models. Working paper.

Chetverikov, D., A. Santos, and A. M. Shaikh (2018). The econometrics of shape restrictions. *Annual Review of Economics*, forthcoming.

Chetverikov, D. and D. Wilhelm (2017). Nonparametric instrumental variable estimation under monotonicity. *Econometrica 85*(4), 1303–1320.

Delecroix, M. and C. Thomas-Agnan (2000). Spline and kernel regression under shape restrictions. In M. G. Schimek (Ed.), *Smoothing and Regression: Approaches, Computation, and Application*, Chapter 5, pp. 109–133. John Wiley & Sons, Inc.

Dette, H., N. Neumeyer, and K. F. Pilz (2006, 06). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli 12*(3), 469–490.

Dierckx, P. (1980). Algorithm/algorithmus 42 an algorithm for cubic spline fitting with convexity constraints. *Computing 24*(4), 349–371.

Donald, S. and H. Paarsch (1993). Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *International Economic Review 34*(1), 121–148.

Du, P., C. F. Parmeter, and J. S. Racine (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica 23*(3), 1347–1371.

Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *The Annals of Statistics 26*(1), 288–314.

Dümbgen, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli 9*(3), 423–449.

Freyberger, J. and J. L. Horowitz (2015). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics 189*(1), 41–53.

Freyberger, J. and Y. Rai (2018). Uniform confidence bands: characterization and optimality. *Journal of Econometrics 204*(1), 119–130.

Freyberger, J. and B. Reeves (2018). Inference under shape restrictions. Working Paper.

Geyer, C. J. (1994). On the asymptotics of constrained m-estimation. *The Annals of Statistics 22*(4), 1993–2010.

Groeneboom, P., G. Jongbloed, and J. A. Wellner (2001). Estimation of a convex function: Characterizations and asymptotic theory. *The Annals of Statistics 29*(6), 1653–1698.

Grundl, S. and Y. Zhu (2015). Identification and estimation of risk aversion in first price auctions with unobserved auction heterogeneity. Working Paper.

Guerre, E. et al. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica 68*(3), 525–574.

Haag, B. R., S. Hoderlein, and K. Pendakur (2009). Testing and imposing slutsky symmetry in nonparametric demand systems. *Journal of Econometrics 153*(1), 33–50.

Haile, P. (2001). Auctions with resale markets: An application to u.s. forest service timber sales. *American Economic Review 91*(3), 399–427.

Haile, P., H. Hong, and M. Schum (2006). Nonparametric tests for common values in first-price sealed-bid auctions. Working Paper.

Haile, P. and E. Tamer (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy 111*(1), 1–15.

Hall, P. and L.-S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics 29*(3), 624–647.

Harstad, R., J. Kagel, and D. Levin (1997). Equilibrium bid functions for auctions with an uncertain number of bidders. *Economics Letters 33*(1), 35–40.

Henderson, D. J. and C. F. Parmeter (2009). Imposing economic constraints in nonparametric regression: Survey, implementation and extension. *Advances in Econometrics 25*, 433–469.

Hendricks, K. and R. Porter (2007). An empirical perspective on auctions. *Handbook of Industrial Organization 3*, 2073–2143.

Hildreth, C. (1954). Point estimates of ordinates of concave functions. *Journal of the American Statistical Association 49*(267), 598–619.

Horowitz, J. L. and S. Lee (2017). Nonparametric estimation and inference under shape restrictions. *Journal of Econometrics 201*(1), 108 – 126.

Jones, D. R. (2001). A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization 21*(4), 345–383.

Kaido, H., F. Molinari, and J. Stoye (2016). Confidence intervals for projections of partially identified parameters. Working paper.

Ketz, P. (2018). Subvector inference when the true parameter vector may be near or at the boundary. *Journal of Econometrics 207*(2), 285–306.

Laffont, J. et al. (1995). Econometrics of first-price auctions. *Econometrica 63*(4), 953–980.

Lewbel, A. (1995). Consistent nonparametric hypothesis tests with an application to slutsky symmetry. *Journal of Econometrics 67*(2), 379–401.

Ma, J. et al. (2018). Inference for first-price auctions with guerre, perrigne, and vuong's estimator. Working Paper.

Mammen, E. (1991a). Estimating a smooth monotone regression function. *The Annals of Statistics 19*(2), 724–740.

Mammen, E. (1991b). Nonparametric regression under qualitative smoothness assumptions. *The Annals of Statistics 19*(2), 741–759.

Mammen, E. and C. Thomas-Agnan (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics 26*(2), 239–252.

Marmer, V. and A. Shneyerov (2012). Quantile-based nonparametric inference for first-price auctions. *Journal of Econometrics 167*(2), 345–357.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Chapter 42, pp. 2524–2558. Elsevier Science.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica 75*(5), 1411–1452.

Mukerjee, H. (1988). Monotone nonparametric regression. *The Annals of Statistics 16*(2), 741–750.

Müller, U. K. and A. Norets (2016). Credibility of confidence sets in nonstandard econometric problems. *Econometrica 84*(6), 2183–2213.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Paarsch, H. (1997). Deriving an estimate of the optimal reserve price : An application to british colombian timber sales. *Journal of Econometrics 78*(2), 337–357.

Pal, J. K. and M. Woodroofe (2007). Large sample properties of shape restricted regression estimators with smoothness adjustments. *Statistica Sinica 17*(4), 1601–1616.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science 3*(4), 425–461.

Riley, J. and W. Samuelson (1981). Optimal auctions. *The American Economic Review 71*(3), 381–392.

Song, U. (2006). Nonparametric identification and estimation of a first-price auction model with an uncertain number of bidders. Working Paper.

Stone, C. (1982). Optimal rates of convergence for nonparametric regressions. *Annals of Statistics 10*(4), 1040–1053.

Wang, X. and J. Shen (2013). Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. *SIAM Journal on Control and Optimization 51*(4), 2753–2787.

Wright, F. T. (1981). The asymptotic behavior of monotone regression estimates. *The Annals of Statistics 9*(2), 443–448.

Zhang, C.-H. (2002). Risk bounds in isotonic regression. *The Annals of Statistics 30*(2), 528–555.