Zhu, Yu

Article

Inference in nonparametric/semiparametric moment equality models with shape restrictions

Quantitative Economics

Provided in Cooperation with:
The Econometric Society

This Version is available at:
http://hdl.handle.net/10419/217197

WWW.ECONSTOR.EU

Mitglied der
Leibniz-Gemeinschaft

Inference in nonparametric/semiparametric moment equality models with shape restrictions

Yu Zhu
Bank of Canada

This paper studies the inference problem of an infinite-dimensional parameter with a shape restriction. This parameter is identified by arbitrarily many unconditional moment equalities. The shape restriction leads to a convex restriction set. I propose a test of the shape restriction, which controls size uniformly and applies to both point-identified and partially identified models. The test can be inverted to construct confidence sets after imposing the shape restriction. Monte Carlo experiments show the finite-sample properties of this method. In an empirical illustration, I apply the method to ascending auctions held by the US Forest Service and show that imposing shape restrictions can significantly improve inference.

Keywords. Nonparametric/semiparametric models, partial identification, shape restrictions, unconditional moments.

JEL classification. C12, C14.

1. Introduction

Economic theories often imply shape restrictions. For example, utility functions are weakly concave in consumption, and demand functions are weakly decreasing in price. These shape restrictions can be useful in empirical studies in two ways. First, they can be used to test competing theories. Second, they provide information that can improve inference on underlying unknown parameters. The latter is especially relevant if the unknown parameters contain functions and a researcher, concerned with misspecification, is unwilling to impose parametric assumptions. In such a case, imposing shape restrictions can lead to significantly smaller confidence sets.

This paper studies an inference problem in which an infinite-dimensional unknown parameter is point-identified or partially identified by unconditional moment equalities. A researcher would like to test whether the unknown parameter satisfies a shape restriction and/or to construct confidence sets after imposing the shape restriction. The unconditional moments may be nonlinear and their sample analog may be discontinuous. The number of the unconditional moments can be finite, countably infinite, or

a continuum. The shape restriction may involve a large number of inequalities and the parameters that satisfy the shape restriction form a convex set. Many commonly used shape restrictions such as non-negativity, monotonicity, concavity, first-order and second-order stochastic dominance satisfy these requirements.

This framework has many applications, and it accommodates conditional and/or unconditional moment equality and/or inequality models. In particular, it accommodates indirect inference in structural estimation. For example, in labor economics, a researcher may want to test whether an individual's reservation wage is decreasing in spells of unemployment (Gutknecht (2016)). In the inference problem of the shape-invariant Engel curve studied in Blundell, Chen, and Kristensen (2007), one may want to test, for example, whether the Engel curve of food is weakly decreasing. Or one may want to impose that the Engel curve is weakly decreasing and construct its confidence set. In first-price auctions, Zhu and Grundl (2014) show that under a weak exclusion condition, inference of bidders' risk aversion is equivalent to testing a first-order stochastic dominance relation. In ascending auctions, one may want to infer valuation distributions with a first-order stochastic dominance restriction that arises from asymmetry in bidders. In these examples, the unknown parameters are identified either by conditional moments (the first two examples), or by a continuum of unconditional moments obtained from indirect inference methods (the last two examples).

In this paper, I propose a test for shape restrictions. It can be inverted to construct confidence sets. The test statistic is obtained by minimizing the integrated squared sample moments subject to the shape restriction. The challenge is that the null distribution of the test statistic depends on a minimization problem taken over an unknown local parameter space, which in turn depends on the identified set of the unknown parameter. A naive estimator of this local parameter space obtained by plugging in an estimator of the identified set may overestimate its size and lead to underestimation of the critical value. To obtain the critical value, I propose a bootstrap procedure that estimates the local parameter space via rescaled sample moments. This method effectively takes a convex combination of 0 and each element in the local parameter space to form an estimator. This estimator is a subset of the original local parameter space if the latter is convex and contains 0, which is in turn guaranteed by the convexity of the restriction set. As a result, convexity of the restriction set ensures that the bootstrap statistic does not over-estimate the size of the local parameter space and validates the test. This method is related to the moment inequality literature where the parameter of interest is finite-dimensional, such as Andrews and Soares (2010), Pakes, Porter, Ho, and Ishii (2015), Andrews and Shi (2013), and, in particular, Gandhi, Lu, and Shi (2013), but the insight is different.

My test applies to both point-identified models and partially identified models, which is a desirable feature because, in many cases, moment conditions only partially identify the unknown parameters.[1] One limitation of the paper is that the theory does

---

[1]For example, Canay, Santos, and Shaikh (2013) showed that in the nonparametric IV regression, the completeness assumption, which is sufficient and necessary for point identification, is not testable. In second-price auction models with incomplete data, the underlying parameter of interest is only partially identified. Finally, in moment inequality models, partial identification arises naturally.

not apply to problems with weak identification. To allow for weak identification is important but more challenging.

In an empirical illustration, I apply my method to infer the marginal valuation distributions of bidders in ascending auctions of timber held by the US Forest Service (USFS). In these auctions, bidders are asymmetric: big firms and firms with manufacturing capacity tend to value timber more than small firms and firms without manufacturing capacity. The valuations of all bidders in an auction are correlated. Because of the auction format, bids other than the transaction price may not correctly reflect bidders' valuations. Partial identification prevails in this setup. Therefore, it fits my method perfectly. Adapting the indirect inference method developed in Bierens and Song (2012) to ascending auctions, I match the model-predicted joint distribution of the transaction price and the winner's identity with the data. This leads to a continuum of unconditional moments that identify the unknown parameter. The key insight is that the asymmetry in bidders implies that the valuation distribution of big firms first-order stochastically dominates (FOSD) that of the small firms. Results show that imposing the FOSD restriction greatly tightens the confidence set and helps to exclude the valuation distribution obtained by incorrectly assuming that bidders are symmetric and have independent valuations. To the best of my knowledge, the inference procedure is new in ascending auctions.

This paper contributes to the growing literature on inference with shape restrictions. It is closely related to Chernozhukov, Newey, and Santos (2015) (CNS hereafter), which provides a general treatment of inference problems with shape restrictions and partial identification. Different from their method, I use a bootstrap statistic that relies on the convexity of the restriction set. As a result, in partially identified models, my test can be faster to compute. Also, unlike CNS, my test can be applied to indirect inference problems with a continuum of moments.

Other recent work in this area includes Chetverikov (2019), Chetverikov and Wilhelm (2017), Fang and Seo (2019), Freyberger and Horowitz (2015), Freyberger and Reeves (2018), Gutknecht (2016), and Horowitz and Lee (2017). All of these papers focus on nonparametric regression or nonparametric IV regression models, except Fang and Seo (2019) and Freyberger and Reeves (2018), which consider general setups but require point identification. This paper also generalizes Hong (2017) and Santos (2012) by allowing for shape restrictions defined by inequalities.

This paper also contributes to the growing literature on semiparametric/nonparametric conditional moment models. Recent work includes Newey, Powell, and Vella (1999), Newey and Powell (2003), Ai and Chen (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011), Chen, Chernozhukov, Lee, and Newey (2014), Chen and Pouzo (2009, 2012, 2015) and Tao (2014).

The paper is organized as follows. Section 2 introduces the econometric framework. Section 3 provides a detailed guide on how to conduct the test and how to obtain confidence sets. Section 4 provides a heuristic illustration of the method. Section 5 lays out the theory. Readers who are less interested in the technical details can skip this section. Section 6 presents Monte Carlo experiments to evaluate the finite-sample performance of the method. Section 7 applies the method to ascending auctions using timber auc-

tions. Section 8 concludes. Technical proofs are collected in the Online Supplemental Material (Zhu (2020)).

2. Setup and examples

The parameter of interest θ_F is infinite-dimensional and lives in a known parameter space Θ. Denote W as a D_W-dimensional random vector, which takes values in $\mathcal{W} \subseteq \mathbb{R}^{D_W}$ and has a distribution F. The unknown parameter θ_F satisfies moment conditions:

$$\mathbb{E}_F \rho_{\mathbf{t}}(W, \theta_F) = 0 \quad \forall \mathbf{t} \in \mathbf{T},$$

where $\rho_{\mathbf{t}}(\cdot, \cdot) : \mathbb{R}^{D_W} \times \Theta \to \mathbb{R}$ is a real function indexed by $\mathbf{t}$, and $\mathbf{T}$ is an index set. $\mathbb{E}_F$ is the expectation taken under the distribution F. The cardinality of $\mathbf{T}$ can be finite, countably infinite, or a continuum. The moment function $\rho_{\mathbf{t}}$ may depend on θ only through its value at W, that is, $\rho_{\mathbf{t}}(W, \theta) = \rho_{\mathbf{t}}(W, \theta(W))$. Or it may depend on the value of some complicated functional applied to θ. The latter often arises in indirect inference of structural models. For an example in first-price auctions, see Bierens and Song (2012).

Even in simple models nested in this framework, partial identification may arise and it is not possible to test partial identification uniformly.[2] It is therefore desirable to have an approach robust to partial identification. To this end, define the identified set Θ_F as the collection of all θs which satisfy the moment conditions under F, that is,

$$\Theta_F = \big\{ \theta \in \Theta : \mathbb{E}_F \rho_{\mathbf{t}}(W, \theta) = 0, \forall \mathbf{t} \in \mathbf{T} \big\}.$$

Notice that Θ_F is allowed to be a singleton, in which case θ_F is point-identified.

A researcher may want to test whether θ_F satisfies some shape restriction or to impose the shape restriction and draw an inference on $\phi(\theta_F)$, where $\phi : \Theta \to \mathbb{R}^{D_\phi}$ is a linear functional and D_ϕ is a positive integer. This paper focuses on the case where the shape restriction is defined by a large number of inequalities. In other words, the shape restriction defines a restriction set

$$R = \big\{ \theta : l(\theta) = a_l, \forall l \in L, \psi(\theta) \geq 0, \forall \psi \in \Psi \big\},$$

where L is a collection of linear functionals, $a_l \in \mathbb{R}$ is a number associated with the linear functional l, and Ψ is a collection of functionals. Both L and Ψ can have finitely or infinitely many elements and R is a convex set. The latter is guaranteed if Ψ contains only convex functionals. To accommodate partial identification, the testing problem is

$$H_0 : \Theta_F \cap R \neq \emptyset,$$

$$H_1 : \Theta_F \cap R = \emptyset.$$

This problem is a generalization of Hong (2017) and Santos (2012), who focus on R defined by equalities of linear operators. A key difference is that if R is defined by inequalities, it has boundary points. Therefore, the inference procedure must account for the possibility that θ_F is close to or at the boundary of R. The formulation of R covers two

---

[2]See Santos (2012) for an example of partial identification in the non-parametric IV case and Canay, Santos, and Shaikh (2013) for an impossibility result on the testability of identification.

inference problems: testing the shape restriction and constructing confidence sets of linear functionals with the shape restriction.

EXAMPLE 2.1 (Nonparametric IV Regression). Consider the model

$$Y = \theta_F(X) + \epsilon,$$

where $\mathbb{E}_F[\epsilon|X] \neq 0$. Let Z be a set of instruments and $\mathbb{E}_F[\epsilon|Z] = 0$. Here, X and Z are D_X- and D_Z-dimensional real random vectors. The goal is to test whether $\theta_F(x)$ is weakly increasing in the first element of x, where x is a D_X-dimensional vector and its kth element is denoted by x_k. Many economic applications fit into this problem. One example is to test whether an individual's reservation wage is decreasing in spells of unemployment, for example, Gutknecht (2016). Assume that θ_F lies in Θ, which is the space of differentiable functions. The identified set is defined by the conditional moment equality

$$\Theta_F = \{\theta \in \Theta : \mathbb{E}_F[Y - \theta(X)|Z] = 0\}.$$

In general, Θ_F might not be a singleton. The restriction set is $R = \{\theta : \partial\theta(x)/\partial x_1 \geq 0, \forall x\}$. It is convex and defined by a continuum of linear functionals: $\psi_x\theta = \partial\theta(x)/\partial x_1, \forall x$. Following Bierens (1990) and Stinchcombe and White (1998), the conditional moment is equivalent to a continuum of unconditional moments with a set of weight functions $w(\cdot, \mathbf{t})$:

$$\Theta_F = \{\theta \in \Theta : \mathbb{E}_F[Y - \theta(X)]w(Z, \mathbf{t}) = 0, \forall \mathbf{t} \in \mathbf{T}\},$$

where $\mathbf{T} \subset \mathbb{R}^{D_Z}$ has a positive Lebesgue measure. Therefore, $W = (X, Y, Z)$ and $\rho_{\mathbf{t}}(W, \theta) = [Y - \theta(X)]w(Z, \mathbf{t})$. One may also want to impose monotonicity to infer $\theta_F(x)$ for some x. Then the linear functional $\phi(\theta) = \theta(x)$. Shape restrictions can greatly help inference in this model. Chetverikov and Wilhelm (2017) showed that monotonicity can improve rates of convergence. If the model is partially identified, shape restrictions can also dramatically reduce the identified set. In extreme cases, moment conditions alone may not provide any information. But with shape restrictions, point identification can be achieved. Section S.1 in the Online Supplemental Material provides such an example. Also notice that the same testing problem but with conditional quantile restrictions also fits into my framework.

EXAMPLE 2.2 (Shape-Invariant Engel Curve). Blundell, Chen, and Kristensen (2007) study semi-nonparametric estimation of the shape-invariant Engel curve. In this estimation problem, a researcher observes $W = (Y_1, Y_2, X, Z)$ from a household, where Y_1 is the fraction of expenditure on a particular good, say food; Y_2 is the total expenditure; X is a vector of household characteristics; and Z is the total income. He is interested in estimating

$$Y_1 = h(Y_2 - X\beta) + \epsilon,$$

where h is the shape-invariant Engel curve and is specified nonparametrically. The total expenditure Y_2 may be endogenous while the total income Z serves as an instrument, that is, $\mathbb{E}_F(\epsilon|Y_2, X) \neq 0$ while $\mathbb{E}_F(\epsilon|Z, X) = 0$. In this case, $\theta = (\beta, h)$ and one can transform the conditional moments into a continuum of unconditional moments as in

the previous example. The researcher may want to test whether the fraction of expenditure on a particular good is decreasing in total expenditure, or may want to conduct an inference on h imposing this decreasing assumption. For example, in an advanced economy like the UK, it is natural to think that the fraction of expenditure on food decreases as total expenditure increases. Assuming that h is differentiable, the restriction set is $R = \{\theta : h'(x) \leq 0, \forall x\}$, which is a convex set.

EXAMPLE 2.3 (Asymmetric Ascending Auctions). In an ascending auction, bidders cry out their bids in ascending order until only one bidder remains. The remaining bidder wins the good and pays the last bid. The auction outcome is equivalent to a sealed bid second-price auction, where the bidder with the highest valuation wins the good and the transaction price is the second highest bid. Consider the model where bidders are asymmetric and have correlated private valuations. Bidders are divided into strong bidders and weak bidders. Let $F_s(v)$ be the marginal valuation distribution of strong bidders and $F_w(v)$ be the marginal valuation distribution of weak bidders. Strong bidders are more likely to have higher valuations for the auctioned object than weak bidders, that is, $F_s(v) \leq F_w(v)$ for every $v \in \mathbb{R}$. Correlation among bidders' valuations is modeled by a Gaussian Copula function. For simplicity, I focus only on auctions with one strong and one weak bidder. Let $\Phi_2(\cdot, \cdot, r)$ be a 2-dimensional normal distribution function with mean 0, variance 1, and correlation r. And let Φ be a standard normal distribution function. Then in an auction, the joint distribution of valuations is $C(F_s(\cdot), F_w(\cdot), r)$, where $C(x_1, x_2, r) = \Phi_2(\Phi^{-1}(x_1), \Phi^{-1}(x_2), r)$. The primitive of this model is $\theta = (F_s, F_w, r)$.

For each auction, a researcher observes the transaction price Y, which is equal to the second highest valuation. He also observes a variable X, which takes value 1 if the winner is a strong bidder, and takes value 0 if otherwise. He is interested in constructing a conference band for F_s and wants to impose the shape restriction $F_s(v) \leq F_w(v)$ for every $v \in \mathbb{R}$ to improve inference.

One can fit this inference problem into my framework using the indirect inference method. The idea is that, at the true parameter, the joint distribution of the transaction price and the winner's type predicted by the model should be close to its empirical counterpart. Following Bierens and Song (2012), one can use characteristic functions to match these two distributions. Then model restrictions reduce to $\forall \mathbf{t} \in \mathbf{T}$,

$$0 = \mathbb{E}_F \sin(\mathbf{t}Y)\mathbf{1}(X = 1) - \mathbb{E}\big[\sin(\mathbf{t}\tilde{Y})\mathbf{1}(\tilde{X} = 1)|\theta\big],$$
$$0 = \mathbb{E}_F \sin(\mathbf{t}Y)\mathbf{1}(X = 0) - \mathbb{E}\big[\sin(\mathbf{t}\tilde{Y})\mathbf{1}(\tilde{X} = 0)|\theta\big],$$
$$0 = \mathbb{E}_F \cos(\mathbf{t}Y)\mathbf{1}(X = 1) - \mathbb{E}\big[\cos(\mathbf{t}\tilde{Y})\mathbf{1}(\tilde{X} = 1)|\theta\big],$$
$$0 = \mathbb{E}_F \cos(\mathbf{t}Y)\mathbf{1}(X = 0) - \mathbb{E}\big[\cos(\mathbf{t}\tilde{Y})\mathbf{1}(\tilde{X} = 0)|\theta\big],$$

where $\mathbf{T}$ is an interval on the real line that contains 0; $\tilde{X}$ and $\tilde{Y}$ are the winner's type and the transaction price generated by the model under parameter θ.[3] The restriction

---

[3]Unlike in Aradillas-López, Gandhi, and Quint (2013) and Coey, Larsen, Sweeney, and Waisman (2017), variation in the number of bidders is not used in this example, and the focus is on the marginal valuation distribution.

set is $R = \{\theta : F_s(v) \le F_w(v) \text{ for all } v \in \mathbb{R}\}$, which is a convex set. It is well known that in this setup, F_s is only partially identified. See Aradillas-López, Gandhi, and Quint (2013), Coey et al. (2017), Komarova (2013). In Section 7, I apply this method to USFS timber auctions.

There are many other applications that fit into my framework. They include the inference of risk aversion in first-price auctions considered in Zhu and Grundl (2014) and Examples 2.2–2.4 in CNS. In particular, my framework also allows for conditional moment inequalities following the strategy used in Examples 2.3 and 2.4 in CNS.

3. Inference procedure

The researcher observes an independently identically distributed (i.i.d.) random sample $\{W_i\}_{i=1}^n$ from F. Denote the probability measure that induces F by $\mathbb{P}_F$.[4]

The test statistic

Let μ be a probability measure on $\mathbf{T}$ such that $\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta) = 0$ for all $\mathbf{t} \in \mathbf{T}$ if and only if

$$Q_F(\theta) = \int_{\mathbf{T}} \left[\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta) \right]^2 d\mu(\mathbf{t}) = 0.$$

Throughout the paper, I assume that such μ exists. If $\mathbf{T}$ contains finitely many or countably infinitely many elements, it is easy to find such a μ. If $\mathbf{T}$ is a closed rectangular in an Euclidean space and $\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)$ is continuous in $\mathbf{t}$, μ can be the uniform probability measure on $\mathbf{T}$.

Ideally, one can compute the infimum (inf) of the criterion $Q_F(\theta)$ on the restriction set R and reject the null hypothesis if it is not 0. Because $Q_F(\theta)$ is unknown, replace it with its sample analog

$$Q_n(\theta) = \int_{\mathbf{T}} \left[\frac{1}{n} \sum_{i=1}^n \rho_{\mathbf{t}}(W_i, \theta) \right]^2 d\mu(\mathbf{t}).$$

The test statistic for the restriction set R is

$$T_n(R) = \inf_{\theta \in \Theta_n \cap R} n Q_n(\theta), \tag{1}$$

where $\{\Theta_n\}_{n=1}^\infty$ is a sequence of increasing sieve spaces that approximate Θ well enough. Here inf is used to allow for potentially discontinuous $Q_n(\theta)$, which arises in, for example, IV quantile regression models. Also define

$$\hat{\Theta}_n(R) = \arg \inf_{\theta \in \Theta_n \cap R} n Q_n(\theta).$$

---

[4]Formally, $\{W_i\}_{i=1}^n$ is the first n coordinate of $\{W_i\}_{i=1}^\infty$, which is the identity mapping on the infinite product probability space $(\mathcal{W}^\infty, \mathcal{S}^\infty, \mathbb{P}_F^\infty)$. Here, $\mathcal{S}^\infty$ is the product σ-algebra and $\mathbb{P}_F^\infty$ is the product measure. This paper relies heavily on convergence of empirical processes. It is understood that if an empirical process is not measurable, convergence is under the outer probability on $(\mathcal{W}^\infty, \mathcal{S}^\infty, \mathbb{P}_F^\infty)$.

A formal definition of $\arg\inf$ is given in Appendix A. At this point, one may regard Q_n as continuous and $\hat{\Theta}_n(R)$ as the set of minimizers. Reject the null hypothesis if $T_n(R)$ exceeds some critical value. Although $T_n(R)$ is based on a CvM-type criterion function, other criterion functions may also be used. One possibility is the Kolmogorov–Smirnov (KS) criterion, which is used in Hong (2017) and Santos (2012). Theories developed in this paper carry naturally over to the case with the KS criterion. This paper focuses on CvM because it is easier to compute.[5]

Bootstrap critical value

This paper develops a bootstrap method to obtain the critical value. Define

$$\mathbb{G}_{n,F}(\theta, \mathbf{t}) = n^{-1/2} \sum_{i=1}^{n} \big[\rho_{\mathbf{t}}(W_i, \theta) - \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta) \big].$$

Let $\{W_i^*\}_{i=1}^n$ be a sequence of i.i.d. draws from the empirical distribution and

$$\mathbb{G}_n^*(\theta, \mathbf{t}) = n^{-1/2} \sum_{i=1}^{n} \bigg[\rho_{\mathbf{t}}(W_i^*, \theta) - \frac{1}{n} \sum_{i=1}^{n} \rho_{\mathbf{t}}(W_i, \theta) \bigg].$$

The bootstrap statistic is defined as

$$T_n^*(R) = \inf_{(\gamma_n, \lambda_n) \in I_n} n Q_n^* \big(\theta_n^*(\gamma_n, \lambda_n, R), \gamma_n, 0 \big), \tag{2}$$

where

$$n Q_n^*(\theta, \gamma_n, \lambda_n) = \int_{\mathbf{T}} \bigg[\mathbb{G}_n^*(\theta, \mathbf{t}) + n^{-1/2} \gamma_n \sum_{i=1}^{n} \rho_{\mathbf{t}}(W_i, \theta) \bigg]^2 d\mu(\mathbf{t}) + \lambda_n Q_n(\theta),$$

$$\theta_n^*(\gamma_n, \lambda_n, R) \in \hat{\Theta}_n^*(\gamma_n, \lambda_n, R) = \arg\inf_{\theta \in \Theta_n \cap R} n Q_n^*(\theta, \gamma_n, \lambda_n),$$

$$I_n \subseteq \overline{I}_n = \big\{ (\gamma, \lambda) \in \mathbb{R}^2 \mid n\gamma^2 + \lambda = \kappa_n, \lambda \geq 0, \gamma \geq 0 \big\}.$$

If $\hat{\Theta}_n^*(\gamma_n, \lambda_n, R)$ contains more than one point, one can assign any of them to $\theta_n^*(\gamma_n, \lambda_n, R)$.[6] Here, I_n can be any nonempty subset of $\bar{I}_n$, and κ_n is a tuning parameter which diverges at a rate slower than n. One may set $\kappa_n = n/\ln n$. Reject the null hypothesis if $T_n(R) > C_n^*(1 - \alpha + \eta, R) + \eta$, where η is an arbitrarily small positive number and $C_n^*(1 - \alpha + \eta, R)$ is the $(1 - \alpha + \eta)$th quantile of the distribution of $T_n^*(R)$ evaluated at

---

[5]If $\rho_{\mathbf{t}}$ is linear in θ, one only needs to solve a quadratic programming problem for many shape restrictions if the criterion function is of the CvM-type. But a KS-type criterion function requires solving a complicated nonlinear optimization problem. In general, the CvM-type criterion tends to have fewer local optimums. This computational difference matters especially because I use a bootstrap method to obtain critical values.

[6]Any point in $\hat{\Theta}_n^*(\gamma_n, \lambda_n, R)$ delivers size control, but different points may lead to power differences. In practice, it suffices to obtain an approximate minimum of $n Q_n^*(\theta, \gamma_n, \lambda_n)$ and then drop the penalty term. I thank the referee for pointing this out.

the empirical distribution. One can set η to a small positive number such as 10^{-8}. Under certain conditions discussed later, it can be set to 0.

Under a proper κ_n, any nonempty $I_n \subseteq \bar{I}_n$ leads to a valid critical value. A natural question is how to pick I_n. This paper recommends using $I_n = \bar{I}_n$ for computationally simple problems and a two-point set $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$ for computationally intensive problems. This is because different values of γ_n and λ_n have different strengths and weaknesses. On the one hand, $\sqrt{n}\gamma_n \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)$ leads to a better approximation of the local parameter space if γ_n is larger. For example, if $(\gamma_n, \lambda_n) = (0, \kappa_n)$, the method uses 0 as a conservative estimate of the local parameter space. The resulting critical value controls size but is conservative asymptotically. If $(\gamma_n, \lambda_n) = (\sqrt{\kappa_n/n}, 0)$, the test has the correct size asymptotically in many cases.[7] On the other hand, estimation error $|n^{-1/2}\gamma_n \sum_{i=1}^n \rho_{\mathbf{t}}(W_i, \theta) - \sqrt{n}\gamma_n \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)|$ is increasing in γ_n, which can worsen the finite-sample performance. Because of this trade-off, it is desirable to have multiple elements in I_n to combine the advantages of different values of (γ_n, λ_n). Naturally, $I_n = \bar{I}_n$ yields the smallest critical value. But simulations suggest that it is important to include $(0, \kappa_n)$ and $(\sqrt{\kappa_n/n}, 0)$. Beyond that, adding more elements does not significantly improve performance. If $(0, \kappa_n) \in I_n$, the resulting critical value remains bounded under any fixed alternative.[8] It is also worth noting that computing the infimum in (2) is a one-dimensional optimization problem if $I_n = \bar{I}_n$, while computing $\theta_n^*(\gamma_n, \lambda_n, R)$ involves an optimization problem which has the same dimension as $\Theta_n \cap R$. In total, computing one bootstrap statistic involves an optimization problem that, at most, has a dimension of 1 plus the dimension of $\Theta_n \cap R$. It reduces further if $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$. This is very convenient for computationally intensive problems. In contrast, the bootstrap statistic proposed in CNS needs to compute an optimization problem which has a dimension twice that of $\Theta_n \cap R$ if partial identification is a concern. The computational gain of my method is nonnegligible if the dimension of $\Theta_n \cap R$ is high and the moments are difficult to evaluate.

Confidence sets

To construct a confidence set for $\phi(\theta_F)$ after imposing some restriction S, let $R_{a,\phi,S} = \{\theta \in S : \phi(\theta) = a\}$ and $\mathbb{K}$ be the parameter space of $\phi(\theta_F)$. A confidence set with $1 - \alpha$ confidence level is

$$\mathcal{C}_n(1 - \alpha) = \{a \in \mathbb{K} : T_n(R_{a,\phi,S}) \le C_n^*(1 - \alpha + \eta, R_{a,\phi,S}) + \eta\}.$$

A step-by-step guide

To construct a test of significance level α:

---

[7]My test is asymptotically nonsimilar, that is, the test does not have the same size uniformly. This is, however, not deficient, as pointed out by Andrews (2012), who shows that asymptotically similar tests have very poor power if there are inequality constraints. If there are no inequality constraints, the test can achieve the exact size asymptotically.

[8]In contrast, the bootstrap methods considered in Santos (2012), Hong (2017), and Gandhi, Lu, and Shi (2013) all lead to diverging critical values under fixed alternatives.

(a) Compute the test statistic $T_n(R)$ according to equation (1).

(b) Draw an i.i.d. sample $\{W_i^*\}_{i=1}^n$ from the empirical distribution F_n and form $\mathbb{G}_n^*(\theta, \mathbf{t})$.

(c) Compute the bootstrap statistic using equation (2). I find $\kappa_n = n/\ln n$ works well in simulations, but theory may rule out this choice in some problems.[9] I recommend $I_n = \bar{I}_n$ if computation is not a concern and $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$ in computationally intensive problems.

(d) Repeat (b)–(c) B times and collect all the bootstrap statistics from each repetition.

(e) Use η plus the $(1 - \alpha + \eta)$th quantile of all the bootstrap statistics as the critical value. Reject the null hypothesis if $T_n(R)$ exceeds this critical value. One may set η to be a very small number, say 10^{-8}.

4. A heuristic illustration

This section heuristically illustrates how the method works and why convexity is crucial using the nonparametric IV setup as in Example 2.1. I focus on the bootstrap statistic with $\lambda_n = 0$, that is, $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$. To abstract from the complications of sieve approximation and partial identification, I further assume that $\Theta_n = \Theta$ and θ_F is point-identified.

By the definition of the test statistic,

$$T_n(R) = \min_{\theta \in \Theta \cap R} \int_{\mathbf{T}} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho_{\mathbf{t}}(W_i, \theta) \right]^2 d\mu(\mathbf{t})$$

$$= \min_{\theta \in \Theta \cap R} \int_{\mathbf{T}} \left\{ \mathbb{G}_{n,F}(\theta, \mathbf{t}) + \sqrt{n} \left[\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta) - \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta_F) \right] \right\}^2 d\mu(\mathbf{t}).$$

Under the null, the minimizer obtained from calculating $T_n(R)$ is close to θ_F for sufficiently large n under appropriate assumptions. If $\mathbb{G}_{n,F}(\theta, \mathbf{t})$ is continuous in θ, its value at this minimizer is close to $\mathbb{G}_{n,F}(\theta_F, \mathbf{t})$. Then use the definition of $\rho_{\mathbf{t}}$ in Example 2.1 and the change of variable $\Delta = \sqrt{n}(\theta - \theta_F)$ to obtain

$$T_n(R) = \min_{\Delta \in \mathbf{V}_n(\theta_F)} \int_{\mathbf{T}} \left\{ \mathbb{G}_{n,F}(\theta_F, \mathbf{t}) - \mathbb{E}_F w(Z_i, \mathbf{t}) \Delta(X_i) \right\}^2 d\mu(\mathbf{t}) + o_{\mathbb{P}_F}(1), \tag{3}$$

where $o_{\mathbb{P}_F}(1)$ is a term that converges to 0 in probability and

$$\mathbf{V}_n(\theta_F) = \{\Delta : \theta_F + \Delta/\sqrt{n} \in \Theta \cap R\}.$$

Here, $-\mathbb{E}_F w(Z, \mathbf{t}) \Delta(X)$ is essentially the directional derivative of $\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)$ at θ_F in the direction Δ. All the directions that need to be evaluated are determined by $\mathbf{V}_n(\theta_F)$, which is the true local parameter space (LPS).

---

[9]Although very desirable, it is difficult to develop a data-driven method to choose κ_n. I instead use simulations to provide guidance on κ_n.

To estimate the null distribution of $T_n(R)$, the main challenge lies in estimating $\mathbf{V}_n(\theta_F)$ because it is, roughly speaking, discontinuous in θ_F. To see this, consider the case where R is the set of nonnegative functions. If $\theta_F = 0$, $\mathbf{V}_n(\theta_F)$ converges to a space that contains only nonnegative functions. If θ_F is strictly positive but arbitrarily close to 0, $\mathbf{V}_n(\theta_F)$ converges to a space that contains all functions. Due to this discontinuity, naively plugging an estimated θ_F into $\mathbf{V}_n(\theta_F)$ can significantly overestimate the size of $\mathbf{V}_n(\theta_F)$, and hence lead to overrejection.

This paper observes that directional derivatives can be approximated by differences between moment conditions evaluated at $\theta_F + \Delta$ and those evaluated at θ_F. Although the true value of θ_F is unknown, $\mathbb{E}_F \rho_\mathbf{t}(W_i, \theta_F)$ is known to be 0. This implies that these differences do not depend on θ_F. If $\gamma_n \to 0$, $\sqrt{n} \frac{\gamma_n}{n} \sum_i \rho_\mathbf{t}(W_i, \theta)$ converges to $\sqrt{n} \gamma_n \mathbb{E}_F \rho_\mathbf{t}(W_i, \theta)$ and

$$
T_n^*(R) = \min_{\theta \in \Theta \cap R} \int_\mathbf{T} \left\{ \mathbb{G}_n^*(\theta, \mathbf{t}) + \sqrt{n} \gamma_n \frac{1}{n} \sum_i \rho_\mathbf{t}(W_i, \theta) \right\}^2 d\mu(\mathbf{t})
$$

$$
= \min_{\theta \in \Theta \cap R} \int_\mathbf{T} \left\{ \mathbb{G}_n^*(\theta, \mathbf{t}) + \sqrt{n} \gamma_n \mathbb{E}_F \rho_\mathbf{t}(W_i, \theta) \right\}^2 d\mu(\mathbf{t}) + o_{\mathbb{P}_F}(1).
$$

If $\sqrt{n} \gamma_n \to \infty$ sufficiently fast, the minimizer obtained from calculating $T_n^*(R)$ is close to θ_F.[10] Then, similar to equation (3) but with $\Delta = \sqrt{n} \gamma_n (\theta - \theta_F)$,

$$
T_n^*(R) = \min_{\Delta \in \mathbf{V}_n^{\gamma_n}(\theta_F)} \int_\mathbf{T} \left\{ \mathbb{G}_n^*(\theta_F, \mathbf{t}) - \mathbb{E}_F w(Z_i, \mathbf{t}) \Delta(X_i) \right\}^2 d\mu(\mathbf{t}) + o_{\mathbb{P}_F}(1), \tag{4}
$$

where my estimated LPS is

$$
\mathbf{V}_n^{\gamma_n}(\theta_F) = \left\{ \Delta : \theta_F + \frac{\Delta}{\sqrt{n} \gamma_n} \in \Theta \cap R \right\}.
$$

Because $\mathbb{G}_n^*(\theta_F, \mathbf{t})$ approximates the distribution of $\mathbb{G}_{n,F}(\theta_F, \mathbf{t})$, $T_n^*(R)$ leads to a valid critical value if $\mathbf{V}_n^{\gamma_n}(\theta_F) \subseteq \mathbf{V}_n(\theta_F)$. This is guaranteed if both Θ and R are convex because $\theta_F \in \Theta \cap R$ and

$$
\theta_F + \frac{\Delta}{\sqrt{n}} = (1 - \gamma_n)\theta_F + \gamma_n \left(\theta_F + \frac{\Delta}{\sqrt{n} \gamma_n} \right).
$$

If there is partial identification and $\rho_\mathbf{t}$ is not linear, a similar argument based on the first-order expansion of the moment conditions goes through.

Figure 1 illustrates how the rescaled moments approximate the true LPS and why convexity is important. In all graphs, $\mathbf{V}_n(\theta_F)/\sqrt{n}$ and $\mathbf{V}_n^{\gamma_n}(\theta_F)/\sqrt{n}$ are defined as the collection of all elements in $\mathbf{V}_n(\theta_F)$ and $\mathbf{V}_n^{\gamma_n}(\theta_F)$ divided by $\sqrt{n}$. Loosely, I will also refer to $\mathbf{V}_n(\theta_F)/\sqrt{n}$ and $\mathbf{V}_n^{\gamma_n}(\theta_F)/\sqrt{n}$ as the true LPS and the estimated LPS in the rest of this section. Figure 1(a) is constructed under $R = \{\theta : \theta(x) \geq 0 \text{ for all } x\}$, which is convex. And

---

[10]The parameter λ_n is introduced to make sure that the minimizers are close to θ_F even if $\sqrt{n} \gamma_n$ is small. This allows me to combine bootstrap statistics with different γ_n. For example, if $\gamma_n = 0$ and there is no λ_n, the minimizers can be far away from θ_F. As a result, size control may fail. See also the discussion following Theorem S.3 in the Online Supplemental Material.
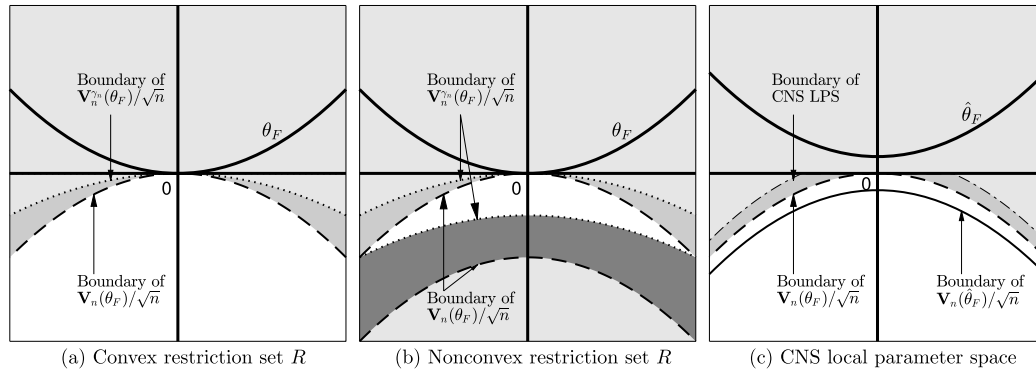
FIGURE 1. Rescaled moments and convexity. Panel (a) depicts the true LPS and the estimated LPS using my method under $R = \{\theta : \theta(x) \geq 0 \text{ for all } x\}$. Panel (b) depicts the true LPS and my LPS under $R = \{\theta : \theta(x) \geq 0 \text{ for all } x \text{ or } \theta(x) \leq -0.5 \text{ for all } x\}$. Panel (c) depicts the true LPS and the estimated LPS in CNS under $R = \{\theta : \theta(x) \geq 0 \text{ for all } x\}$.

θ_F is the black solid line that lies above the x-axis, and hence lies in R. The true LPS and the estimated LPS are the areas above their boundaries, respectively. Both spaces are convex because if two functions lie above one of the two boundaries, their convex combinations also lie above that boundary. The estimated LPS is obtained by shrinking the true LPS toward 0 by a factor of γ_n. In other words, it is obtained by taking a convex combination of 0 and each element in the true LPS. The light gray area is shared by the true LPS and the estimated LPS, while the medium gray area is in the true LPS but not in the estimated LPS. Because of the convexity, the estimated LPS is contained in the true LPS. This implies that $T_n(R)$ is obtained by a minimization problem over a larger set compared to $T_n^*(R)$. Therefore, the distribution of $T_n(R)$ is first-order stochastically dominated by that of $T_n^*(R)$ asymptotically, which validates the bootstrap critical value. Notice that the medium gray area is smaller if γ_n is larger, which means that a large γ_n yields a better approximation of the true LPS for any given n. It is also worth noting that even though γ_n converges to 0, the limit of $\mathbf{V}_n^{\gamma_n}(\theta_F)$ can be the same as that of $\mathbf{V}_n(\theta_F)$ as $n \to \infty$.

Figure 1(b) is constructed under $R = \{\theta : \theta(x) \geq 0 \text{ for all } x \text{ or } \theta(x) \leq -0.5 \text{ for all } x\}$, which is not convex. Now the true LPS has two boundary lines: one passes through the origin and the other lies below the x-axis. It consists of functions that lie above the higher boundary and functions that lie below the lower boundary. Obviously, it is not convex. Again, the estimated LPS is obtained by shrinking the true LPS toward 0. As in Figure 1(a), the light gray region is shared by both the true LPS and the estimated LPS, and the medium gray region lies in the true LPS but not in the estimated LPS. But there is now a dark gray area, which lies in the estimated LPS but not in the true LPS. Therefore, $T_n(R)$ is no longer obtained by a minimization problem over a larger region compared to $T_n^*(R)$. Hence, the size control may not hold. Such differences can matter asymptotically and one can construct examples where uniform size control fails due to lack of convexity.

Lastly, it is useful to compare my method with CNS. The light gray region in Figure 1(c) corresponds to the estimator of LPS in CNS under $R = \{\theta : \theta(x) \geq 0 \text{ for all } x\}$. It is obtained by shifting the boundary of $\mathbf{V}_n(\hat{\theta}_F)/\sqrt{n}$ toward 0, where $\hat{\theta}_F$ is a constrained estimator of θ_F. The size of the shift at a point x is the minimum of a tuning parameter and the distance from $\hat{\theta}_F(x)$ to the boundary of the restriction set. This method differs from mine in two ways. First, because the rescaled moments exploit the fact that the moments are 0 at θ_F, my estimated LPS is not explicitly based on an estimator of θ_F. Second, my method shrinks the true LPS toward 0 by a factor of γ_n. Compared to Figure 1(a), the medium gray region in Figure 1(c), which is the set of function values that are in the true LPS but not in the LPS of CNS, is wider around the origin but gets narrower away from the origin. This highlights the difference between the two methods.

5. Asymptotic theory

This section provides a theoretic justification of my test. To simplify the presentation, the main assumptions and related technical discussions are collected in Appendix B. Intuitively, these assumptions have the following important implications. First, they rule out weak identification. Second, $\Theta_n \cap R$ should be a suitable approximation for $\Theta \cap R$. In particular, the approximation error needs to be sufficiently small. Third, if the moment conditions are nonlinear, the rate of convergence of $\hat{\Theta}_n(R)$ needs to be sufficiently fast. If this rate is faster, κ_n is allowed to diverge at a slower rate. If this convergence rate is too slow, a valid choice for κ_n may not exist.

In the rest of the paper, I assume that Θ is a convex space. The theoretic results of this paper focus on the following class of distribution functions.

DEFINITION 5.1. Let $\mathcal{F}$ be a collection of distribution functions F such that

(i) The moment function $\rho_{\mathbf{t}}(\cdot, \theta)$ is measurable for all $(\theta, \mathbf{t}) \in \Theta \times \mathbf{T}$.

(ii) There exists a measurable function $\mathbf{F}$ such that $|\rho_{\mathbf{t}}(\cdot, \theta)| \leq \mathbf{F}(\cdot)$ for all $(\theta, \mathbf{t}) \in \Theta \times \mathbf{T}$, and $\sup_{F \in \mathcal{F}} \mathbb{E}_F \mathbf{F}(W_i)^2 < \infty$.

(iii) The set of functions $\varrho = \{\rho_{\mathbf{t}}(\cdot, \theta) : (\theta, \mathbf{t}) \in \Theta \times \mathbf{T}\}$ is Donsker and pre-Gaussian uniformly in $F \in \mathcal{F}$.

Part (i) in Definition 5.1 is a mild condition and is automatically satisfied if $\rho_{\mathbf{t}}(W_i, \theta)$ is continuous in W_i for all $(\theta, \mathbf{t}) \in \Theta \times \mathbf{T}$. Part (ii) requires that there is an envelope function for $\rho_{\mathbf{t}}(\cdot, \theta)$. And the envelope function has a uniformly bounded second moment. Part (iii) is the key to the uniform convergence rates and the validity of the test. Under the null hypothesis, the test statistic can be expressed as a continuous function applied to $\mathbb{G}_{n,F}$, and the bootstrap statistic can be expressed as the same function applied to $\mathbb{G}_n^*$. Part (iii) then allows me to invoke a result in Linton, Song, and Whang (2010) to establish uniform size control. This condition can be satisfied after regularizing the parameter space, following Newey and Powell (2003) and Santos (2012). Section S.2 in the Online Supplemental Material provides low-level conditions that guarantee part (iii).

DEFINITION 5.2. Let $\mathcal{J} = \{(F, R) : F \in \mathcal{F}, R \in \mathcal{R}, \text{ and } \Theta_F \cap R \neq \emptyset\}$, where $\mathcal{R}$ is a set of nonempty, convex sets.

The set $\mathcal{J}$ consists of pairs of the distribution function and restriction set that satisfy the null hypothesis. I now introduce the main result of this paper, which shows that the proposed test controls size uniformly on $\mathcal{J}$.

THEOREM 5.1. *Under Assumptions* B.1–B.5, *for any nonempty* $I_n \subseteq \bar{I}_n$ *and* $\eta > 0$:

(i) $\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J}} \mathbb{P}_F(T_n(R) > C_n^*(1 - \alpha + \eta, R) + \eta) \leq \alpha$. *If in addition Assumption* B.7 *holds, this inequality holds with* $\eta = 0$.

(ii) *If* $F \in \mathcal{F}$, $\Theta_F \cap R = \emptyset$, *then* $\mathbb{P}_F$ *almost surely,* $T_n(R) > C_n^*(1 - \alpha + \eta, R) + \eta$ *as* $n \to \infty$.

Theorem 5.1 establishes the validity of the test. Here, η appears for technical reasons. Under Assumption B.7, which consists of an anticoncentration condition and a condition that handles the degenerate case, η can then be set to 0.

Now I briefly discuss the assumptions. Assumption B.1 requires that $\Theta \cap R$ is compact, and that $\mathbb{E}_F \rho_t(W, \theta)$ is continuous in θ. It guarantees that the test statistic diverges to infinity under a fixed alternative, which ensures the second claim of Theorem 5.1. Assumption B.2 requires that the sieve spaces approximate Θ well such that the approximation bias can be ignored. Assumption B.3 is necessary to obtain the convergence rates of $\hat{\Theta}_n(R)$ and $\hat{\Theta}_n^*(\gamma_n, \lambda_n, R)$. These rates are important for size control. Assumption B.4 serves two roles. First, combined with Definition 5.1(iii), it implies that $\mathbb{G}_{n,F}$ and $\mathbb{G}_n^*$ are asymptotically equi-continuous in θ. This allows us to replace θ by θ_F in equations (3) and (4). Second, it guarantees that the moment conditions can be approximated by their linear expansions. Therefore, the heuristic argument in Section 4 based on linearity can be extended to nonlinear moments. Lastly, Assumption B.5 puts restrictions on κ_n. It implicitly rules out some slow convergence rate including the severely ill-posed problems.

REMARK 5.1. Assumptions B.4(i), (ii), and the convexity of Θ and R are needed to use a general I_n. If $I_n = \{(0, \kappa_n)\}$, my test remains valid without these assumptions. Then Assumption B.5 can also be relaxed, that is, κ_n can be chosen to be any diverging sequence that does not diverge too fast. In addition, one does not need restrictions on the rates of convergence of $\hat{\Theta}_n(R)$. However, the resulting test can be very conservative. As illustrated in Section 4, a lower γ_n yields a potentially more conservative approximation. If $I_n = \{(0, \kappa_n)\}$, γ_n is set to be 0, which is the smallest value.

REMARK 5.2. An immediate consequence of Theorem 5.1 is the validity of $\mathcal{C}_n(1 - \alpha)$. To see this, just notice that one can set $\mathcal{R} = \{R_{a,\phi,S} : a \in \mathbb{K}\}$ and

$$\liminf_{n \to \infty} \inf_{F \in \mathcal{F} : \theta_F \in S} \mathbb{P}_F(\phi(\theta_F) \in \mathcal{C}_n(1 - \alpha))$$

$$\geq \liminf_{n \to \infty} \inf_{F \in \mathcal{F} : \theta_F \in S} \mathbb{P}_F(T_n(R_{\phi(\theta_F),\phi,S}) \leq C_n^*(1 - \alpha + \eta, R_{\phi(\theta_F),\phi,S}) + \eta)$$

$$\geq \liminf_{n \to \infty} \inf_{(F,R) \in \mathcal{J}} \mathbb{P}_F(T_n(R) \leq C_n^*(1 - \alpha + \eta, R) + \eta) \geq 1 - \alpha.$$

COROLLARY 5.1. *Under Assumptions* B.1–B.6,

$$\liminf_{n \to \infty} \inf_{F \in \mathcal{F} : \theta_F \in S} \mathbb{P}_F\big(\phi(\theta_F) \in \mathcal{C}_n(1 - \alpha)\big) \geq 1 - \alpha.$$

REMARK 5.3. To make the bootstrap procedure even faster, one might also consider bootstrapping

$$\dot{T}_n^*(R) = \inf_{\theta \in \Theta_n \cap R \cap \hat{\Theta}_n^\epsilon(R)} \int_{\mathbf{T}} \left[\mathbb{G}_n^*(\theta, \mathbf{t}) + \frac{\gamma_n}{\sqrt{n}} \sum_{i=1}^n \rho_{\mathbf{t}}(W_i, \theta) \right]^2 d\mu(\mathbf{t}),$$

where $\epsilon > 0$ is a constant and $\hat{\Theta}_n^\epsilon(R)$ is the ϵ-neighborhood of $\hat{\Theta}_n(R)$. One can prove that this bootstrap is valid following the same method used in this paper.

REMARK 5.4. In general, Theorem 5.1(i) holds with weak inequality. But if the shape restriction does not involve inequality constraints, it can be shown that the test achieves the exact size asymptotically. In addition, my test cannot be ranked with CNS in terms of power as neither dominates the other. This can be seen by comparing Figures 1(a) and 1(c). There are functions that lie in the light gray region in Figure 1(a) but not in the light gray region in Figure 1(c) and vice versa. I also provide an example on the power comparison in the Online Supplemental Material.

6. MONTE CARLO EXPERIMENTS

This section considers three data generating processes (DGPs). The first two DGPs follow Example 2.1, where critical values are relatively easy to calculate. This allows me to investigate extensively how the performance of the test depends on the parameter space, I_n, κ_n, and the approximation of the integration. The last DGP follows Example 2.3, where critical values are more difficult to calculate. I use it to illustrate how the test works in an environment that is close to my empirical application and to provide some guidance on the choice of κ_n.

6.1 *Monotonicity in nonparametric IV regression*

Now consider Example 2.1 with the following two DGPs.

DGP I This DGP is considered in Santos (2012). Generate a latent i.i.d. sample by

$$\begin{pmatrix} X^* \\ Z^* \\ \epsilon^* \end{pmatrix} \sim N \left(0, \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix} \right).$$

Construct X, Z, ϵ by setting

$$X = 2\big(\Phi(X^*/3) - 0.5\big), \qquad Z = 2\big(\Phi(Z^*/3) - 0.5\big), \qquad \epsilon = \epsilon^*,$$

where Φ is the distribution function of a standard normal random variable.
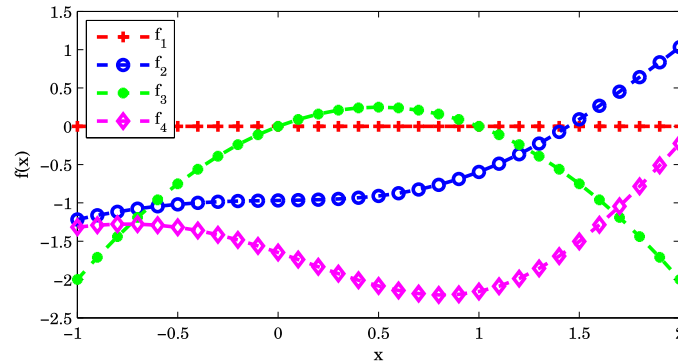
FIGURE 2. Regression functions.

DGP II $Z \sim \text{Uniform}[0, 1]$ and $X = Z + U$ where $U \sim \text{Uniform}[-1, 1]$ and is independent of Z. In addition, $\epsilon = U + 0.1\epsilon^*$ where $\epsilon^* \sim N(0, 1)$ and is independent of (U, X, Z). Newey, Powell, and Vella (1999) show that θ_F is point-identified up to a constant with the additional moment condition $\mathbb{E}_F(U|Z) = 0$. I use only the moment condition $\mathbb{E}_F(\epsilon|Z) = 0$. Then the model is in general partially identified but can be point-identified with shape restrictions in some cases. For more details, see Section S.1 in the Online Supplemental Material.

For both DGPs, I consider $\theta_F = f_i$, $i = 1, 2, 3, 4$ where $f_1(X) = 0$, $f_2(X) = X - 4\varphi(X - 1)$, $f_3(X) = X(1 - X)$, and $f_4(X) = X - 8 \times \varphi(X - 1.15)$. Here, φ is the standard normal density function. These functions are plotted in Figure 2. f_1 is a constant function. f_2 is strictly increasing, while f_3 and f_4 are decreasing in some regions. However, because this model may be partially identified, data generated by f_3 and f_4 may still satisfy the null hypothesis. In fact, in DGP I with $\theta_F = f_3$, the null hypothesis is satisfied. In DGP II, only f_1 and f_2 satisfy the null hypothesis.

I generate i.i.d. samples with 1000 observations. The sieve space is chosen to be spanned by B-splines. In DGP I, I use knots $\{-1, -1, -\frac{1}{3}, \frac{1}{3}, 1, 1\}$ to approximate f_1, f_2, and f_3, and knots $\{-1, -1, -\frac{3}{4}, -\frac{1}{2}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 1\}$ for f_4. In model II, I use knots $\{-1, -1, 0, 1, 2, 2\}$ for f_1, f_2, f_3, and

$$\{-1, -1, -0.625, -0.25, 0.125, 0.5, 0.875, 1.25, 1.625, 2, 2\}$$

for f_4. Different knots are used because X is supported on $[-1, 1]$ in DGP I and on $[-1, 2]$ in DGP II. I also use more knots for f_4 because it is harder to approximate. The weight function is defined as $w(z, \mathbf{t}) = \varphi(\frac{t_1 - z}{t_2})$ and $\mathbf{t} = (t_1, t_2) \in \mathbf{T}$ with $\mathbf{T} = [-0.8, 0.8] \times [0.05, 0.2]$. This weight function is also used in Santos (2012). The measure μ is chosen to be the uniform probability measure on $\mathbf{T}$. I approximate the numerical integration by averaging over points in $\mathbf{T}$. To evaluate the sensitivity of the test to the numerical integration, I consider two cases. In the first case, I average over 14 points, which is constructed by taking the product of 7 equally spaced points on $[-0.8, 0.8]$ and 2 equally spaced points on $[0.05, 0.2]$. In the second case, I average over 350 points obtained by taking the product of 35 equally spaced points on $[-0.8, 0.8]$ and 10 equally spaced points on

[0.05, 0.2]. This effectively uses 350 moments to construct the test. I restrict the parameter space such that the B-spline coefficients are no larger than **B** in absolute values. The value of **B** ranges from 5 to 100 to evaluate its influence.

I consider the following choices of I_n: (1) $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$. This leads to a test that is in spirit similar to Gandhi, Lu, and Shi (2013) and is denoted by GLS. (2) $I_n = \{(0, \kappa_n)\}$, which is otherwise the same as in Hong (2017) except it is based on a CvM criterion function and does not include the penalty term. Since this choice uses the most conservative estimate for the local parameter space, I denote it as least favorable (LF). (3) I also consider bootstrap statistics with

$$I_n = \left\{ (\gamma, \lambda) \,\middle|\, n\gamma^2 + \lambda = \kappa_n, \lambda \geq 0, \gamma \geq 0, \gamma = \frac{i}{\bar{N}}\sqrt{\frac{\kappa_n}{n}}, i = 0, 1, 2, \ldots, \bar{N} \right\}.$$

Each of these bootstrap statistics is the minimum of several bootstrap statistics with different (γ_n, λ_n), and hence is called the Minimum Bootstrap statistic (MB). To assess the sensitivity of the method to I_n, I consider tests with $\bar{N} = 1$ (MB1) and $\bar{N} = 30$ (MB2). The performance of the test can be sensitive to κ_n. Therefore, I report results for $\kappa_n = n^{2/3}, n/\ln n$ to evaluate its impact. Critical values are computed with 500 bootstrap statistics and η is set to 0. Setting $\eta = 10^{-8}$ yields identical results. I simulate 1000 random samples to compute the rejection probability. The significance level is set to 10%.

Table 1 summarizes the rejection probability on the simulated data. The first column shows the true θ_F. The other columns report rejection probabilities of corresponding tests.

First, let us focus on the size control. Under DGP I, the rejection probabilities should be no larger than 10% for f_1, f_2, and f_3, while under DGP II, the rejection probabilities should be no larger than 10% for f_1 and f_2. There are several interesting findings. First, the MBs perform very well. If $\kappa_n = n/\ln n$, they have the right size control in both DGPs. In particular, with f_1 under DGP II, the rejection probabilities are very close to the nominal size 10%. While under DGP I with f_1, the rejection probability is also not far from 10%. This is mostly true with $\kappa_n = n^{2/3}$ except that in some cases, there is slight over-rejection. Second, MB1 performs significantly better than LF and GLS. The latter two can be very conservative. This highlights the importance of including multiple points in I_n. Third, MB1 performs similarly to MB2 although it has much fewer points in I_n. This suggests that the method is not sensitive to the choice of I_n as long as both $(0, \kappa_n)$ and $(\sqrt{\kappa_n/n}, 0)$ are included. Fourth, the test does not seem to be sensitive to the choices of **B** and the number of moments in terms of size control.

Now let us turn to power. First, notice that GLS and LF have much lower power compared to MB1 and MB2. The power difference can be more than 20%. This highlights the power gain by an I_n with multiple elements. Second, MB1 and MB2 have almost identical performance. This is good news for computationally intensive problems because one can set $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$ without losing much. Third, **B** has a limited impact on the results. Fourth, using more moments may decrease (against f_3 in DGP II) or increase (against f_4 in DGP II) power in finite samples. This is not surprising. Intuitively, if one includes informative moments, power increases. Otherwise, power can decrease. Lastly, a smaller κ_n increases the power of the test in certain cases. This is because a smaller κ_n

TABLE 1. Rejection probability with a 10% significance level under $n = 1000$. GLS: $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$; LF: $I_n = \{(0, \kappa_n)\}$; MB1: $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$; and MB2: I_n contains 31 equally spaced points between $(0, \kappa_n)$ and $(\sqrt{\kappa_n/n}, 0)$. The number of moments corresponds to how many points are used to approximate the integral. And **B** is the bound imposed on the parameter space.

θ_F	$\kappa_n = n/\ln n$								$\kappa_n = n^{2/3}$							
	GLS	LF	MB1	MB2	GLS	LF	MB1	MB2	GLS	LF	MB1	MB2	GLS	LF	MB1	MB2
DGP I
	\multicolumn{4}{c}{**B** = 5, 14 moments}	\multicolumn{4}{c}{**B** = 5, 350 moments}	\multicolumn{4}{c}{**B** = 5, 14 moments}	\multicolumn{4}{c}{**B** = 5, 350 moments}
f_1	0.017	0.002	0.056	0.058	0.028	0.006	0.068	0.069	0.042	0.003	0.072	0.076	0.050	0.006	0.079	0.081
f_2	0.025	0.002	0.049	0.050	0.016	0.004	0.040	0.043	0.026	0.000	0.039	0.039	0.027	0.003	0.060	0.061
f_3	0.011	0.000	0.018	0.020	0.006	0.000	0.020	0.020	0.014	0.001	0.020	0.020	0.015	0.000	0.023	0.023
f_4	0.812	0.640	0.909	0.914	0.785	0.670	0.912	0.914	0.847	0.674	0.918	0.919	0.863	0.706	0.937	0.937
	\multicolumn{4}{c}{**B** = 10, 14 moments}	\multicolumn{4}{c}{**B** = 10, 350 moments}	\multicolumn{4}{c}{**B** = 10, 14 moments}	\multicolumn{4}{c}{**B** = 10, 350 moments}
f_1	0.023	0.005	0.068	0.071	0.023	0.008	0.064	0.067	0.033	0.002	0.072	0.077	0.041	0.010	0.082	0.083
f_2	0.013	0.002	0.038	0.039	0.017	0.001	0.041	0.044	0.028	0.004	0.063	0.063	0.033	0.008	0.065	0.065
f_3	0.013	0.000	0.022	0.022	0.011	0.000	0.019	0.020	0.011	0.000	0.020	0.020	0.007	0.000	0.016	0.018
f_4	0.814	0.658	0.911	0.915	0.800	0.681	0.912	0.916	0.859	0.662	0.922	0.924	0.839	0.708	0.916	0.917
	\multicolumn{4}{c}{**B** = 100, 14 moments}	\multicolumn{4}{c}{**B** = 100, 350 moments}	\multicolumn{4}{c}{**B** = 100, 14 moments}	\multicolumn{4}{c}{**B** = 100, 350 moments}
f_1	0.031	0.004	0.075	0.076	0.023	0.011	0.079	0.082	0.042	0.003	0.085	0.086	0.054	0.015	0.090	0.095
f_2	0.023	0.001	0.057	0.058	0.016	0.004	0.055	0.057	0.038	0.004	0.067	0.067	0.032	0.003	0.064	0.065
f_3	0.017	0.000	0.022	0.022	0.008	0.000	0.011	0.013	0.015	0.000	0.026	0.026	0.012	0.000	0.017	0.017
f_4	0.766	0.583	0.935	0.940	0.723	0.599	0.900	0.913	0.858	0.680	0.943	0.947	0.857	0.707	0.937	0.944
DGP II
	\multicolumn{4}{c}{**B** = 5, 14 moments}	\multicolumn{4}{c}{**B** = 5, 350 moments}	\multicolumn{4}{c}{**B** = 5, 14 moments}	\multicolumn{4}{c}{**B** = 5, 350 moments}
f_1	0.027	0.031	0.089	0.091	0.031	0.041	0.100	0.101	0.032	0.044	0.113	0.116	0.045	0.074	0.136	0.137
f_2	0.009	0.000	0.018	0.019	0.008	0.002	0.017	0.017	0.012	0.000	0.023	0.025	0.013	0.001	0.026	0.028
f_3	0.447	0.293	0.592	0.593	0.424	0.314	0.549	0.557	0.496	0.424	0.662	0.675	0.419	0.372	0.561	0.569
f_4	0.494	0.483	0.748	0.749	0.656	0.685	0.857	0.860	0.579	0.606	0.799	0.803	0.743	0.777	0.904	0.905
	\multicolumn{4}{c}{**B** = 10, 14 moments}	\multicolumn{4}{c}{**B** = 10, 350 moments}	\multicolumn{4}{c}{**B** = 10, 14 moments}	\multicolumn{4}{c}{**B** = 10, 350 moments}
f_1	0.024	0.026	0.080	0.083	0.030	0.043	0.096	0.099	0.041	0.058	0.131	0.136	0.044	0.073	0.136	0.139
f_2	0.007	0.002	0.020	0.022	0.006	0.002	0.022	0.023	0.013	0.004	0.027	0.027	0.009	0.004	0.026	0.027
f_3	0.457	0.302	0.605	0.613	0.421	0.304	0.573	0.583	0.497	0.420	0.666	0.676	0.425	0.398	0.592	0.601
f_4	0.558	0.527	0.784	0.791	0.680	0.694	0.854	0.856	0.624	0.632	0.829	0.836	0.763	0.795	0.898	0.898
	\multicolumn{4}{c}{**B** = 100, 14 moments}	\multicolumn{4}{c}{**B** = 100, 350 moments}	\multicolumn{4}{c}{**B** = 100, 14 moments}	\multicolumn{4}{c}{**B** = 100, 350 moments}
f_1	0.024	0.026	0.086	0.088	0.027	0.037	0.086	0.092	0.030	0.044	0.110	0.112	0.043	0.072	0.144	0.145
f_2	0.005	0.000	0.016	0.017	0.007	0.000	0.018	0.018	0.007	0.001	0.015	0.015	0.011	0.002	0.024	0.025
f_3	0.452	0.313	0.592	0.596	0.397	0.288	0.528	0.535	0.498	0.425	0.671	0.678	0.450	0.405	0.594	0.601
f_4	0.542	0.506	0.760	0.768	0.713	0.699	0.879	0.880	0.691	0.654	0.851	0.854	0.774	0.771	0.901	0.906

increases the size of the imposed identified set, and hence can lead to smaller critical values. But it is worth noting that this can also lead to overrejection under the null.

6.2 *Asymmetric ascending auctions*

DGP III   Now consider Example 2.3 where F_s and F_w are beta distributions with parameters $(2, 3)$ and $(2, 3.5)$, respectively. The true value of r in the Gaussian Copula function is 0.8. Each auction has two bidders. One of them is strong and the other is weak. This DGP closely follows the empirical application. The goal here is to construct a confidence interval for F_s evaluated at 0.5, whose true value is 0.6875.

I generate 500 auctions and use B-splines with knots $\{0, 0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1, 1\}$ to approximate F_s and F_w. Because F_s and F_w are distribution functions, I impose that they are weakly increasing and $F_s(0) = F_w(0) = 0$, $F_s(1) = F_w(1) = 1$. I impose the FOSD restriction, that is, $F_s(v) \le F_w(v)$ for all v. I also impose that bidders' valuations are positively correlated and that the correlation is not too low, that is, $r \ge 0.5$.[11]

Given θ, theoretic moments predicted by the model involve integrals that do not have closed forms. I approximate the theoretic moments by Monte Carlo integration with 30,000 simulated auctions. The measure μ is chosen to be the uniform probability measure on $\mathbf{T} = [-1, 1]$. I approximate the integrals by averaging over 100 equally spaced points in $[-1, 1]$. For the bootstrap statistic, I consider GLS, LF, and MB1 as in the previous subsection. I report MB3 with $I_n = \{(0, n/\ln n), (\sqrt{1/\ln n}, 0), (\sqrt{1/2\ln n}, n/2\ln n)\}$ instead of MB2 because of computational difficulties. As in Section 6.1, I restrict the coefficients on the B-splines to be no larger than $\mathbf{B}$ in absolute values. I only consider $\mathbf{B} = 5$ and $\mathbf{B} = 100$. Again, $\eta = 0$ and $\eta = 10^{-8}$ lead to identical rejection probabilities. To assess the sensitivity of my method to κ_n, I consider $\kappa_n = n^{2/3}$ and $n/\ln n$. Critical values are computed based on 200 bootstrap statistics and results are obtained from 500 independent samples. The nominal size of these tests is 10%.

Figure 3 reports rejection probabilities against different values of $F_s(0.5)$. The first row is constructed under $\mathbf{B} = 5$. If $\kappa_n = n/\ln n$ as shown in Figure 3(a), all four tests have correct size control. Similar to the results from DGP I and DGP II, MB1 and MB3 are less conservative than GLS and LF under the null and have much better power. Including both $(0, \kappa_n)$ and $(\sqrt{\kappa_n/n}, 0)$ in I_n significantly improves power. Interestingly, LF seems to be very conservative and does not have any power against alternatives larger than the true value. Also, notice that MB1 and MB3 have identical rejection probabilities, that is, adding $(\sqrt{1/2\ln n}, n/2\ln n)$ does not seem to affect the performance of the test. Figure 3(b) is constructed under $\kappa_n = n^{2/3}$. It is very similar to Figure 3(a). Figure 3(c) compares the rejection probabilities of MB1 under $\kappa_n = n/\ln n$ and $\kappa_n = n^{2/3}$. The performance of the test is not sensitive to κ_n under this DGP. Overall, the test seems to perform slightly better with $\kappa_n = n/\ln n$. This finding is suggestive on the choice of κ_n in Section 7 because the application is very close to DGP III.

The second row of Figure 3 is constructed with $\mathbf{B} = 100$. It is similar to the first row. In particular, the rejection probabilities are very close to those under $\mathbf{B} = 5$, and the performance of the tests are not sensitive to the choice of κ_n.

It is worth noting that accounting for partial identification is desirable here because the model is only partially identified. At the same time, the dimension of the optimization problem is relatively high (17 parameters) and the moment conditions are nonlinear and difficult to evaluate. Solving an optimization problem with a dimension twice that of the sieve space can be very time consuming. For such applications, the method proposed in this paper can be very attractive.

To sum up, all the simulation results suggest that MB1 works very well. Adding more points to I_n only slightly improves the performance of the test. In addition, $\kappa_n = n/\ln n$ seems to be a good choice and results are not sensitive to the value of $\mathbf{B}$.

---

[11]Fan, He, and Li (2015) also use such parametric restrictions on copula functions to partially identify primitives in auctions with correlated values.
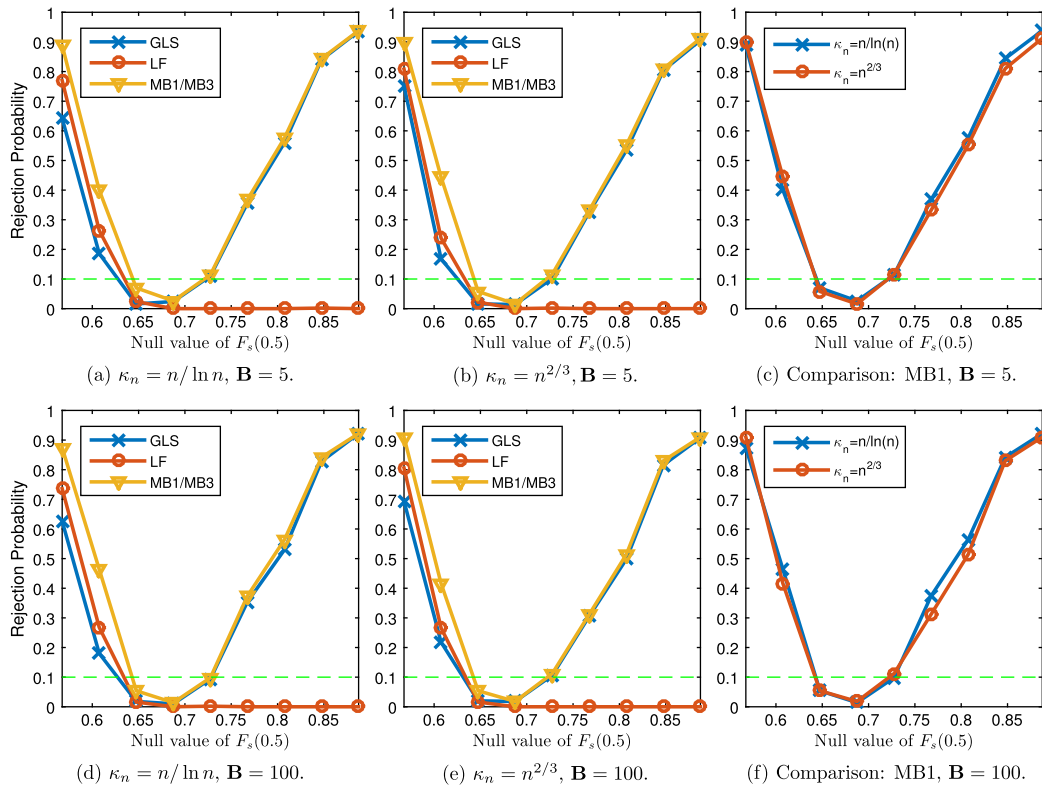
FIGURE 3. Rejection probability with a 10% significance level test. GLS: $I_n = \{(\sqrt{\kappa_n/n}, 0)\}$; LF: $I_n = \{(0, \kappa_n)\}$; MB1: $I_n = \{(0, \kappa_n), (\sqrt{\kappa_n/n}, 0)\}$; and MB3: $I_n = \{(0, n/\ln n), (\sqrt{1/\ln n}, 0), (\sqrt{1/2\ln n}, n/2\ln n)\}$. **B** is the bound imposed on the parameter space. The true value is 0.6875.

7. Empirical application: USFS ascending auctions

This section applies my method to infer marginal valuation distributions in ascending auctions held by the USFS. In these auctions, bidders are timber firms and the auctioned goods are timber tracts. Bidders in these auctions are asymmetric: big firms and firms with manufacturing capacity are more likely to have higher values compared to small firms and firms without manufacturing capacity. The former are thus strong bidders and the latter are weak bidders. For a discussion on the asymmetry, see Athey, Levin, and Seira (2011). In addition, these bidders have correlated valuations, and ignoring this correlation can lead to inconsistent inference. For example, see Aradillas-López, Gandhi, and Quint (2013) and Aradillas-López, Gandhi, and Quint (2016). Therefore, the econometric framework described in Example 2.3 is appropriate for this analysis.

I focus on auctions from region 5 of USFS that are not set-aside auctions. From each auction, I observe every bidder's identity, all the bids, and a set of auction characteristics. I classify bidders as strong or weak according to the SBA classification. To simplify the analysis, I use auctions with 1 strong bidder and 1 weak bidder. To eliminate outliers, I exclude auctions with the highest 5% of transaction prices. This gives me a sample with 218 observations. The average transaction price is $51.13 with a standard devia-
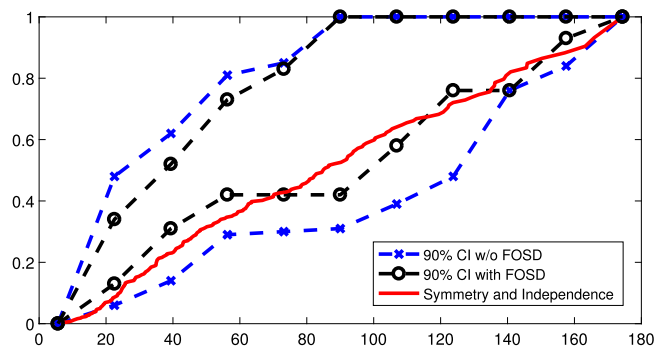
FIGURE 4. Pointwise confidence set for F_s.

tion of \$38.82. Following Aradillas-López, Gandhi, and Quint (2013), I assume that the transaction price equals the second highest valuation among all bidders. To keep the illustration simple, I do not include any auction-level characteristics.

The goal is to construct a pointwise confidence band for F_s. I implement my method as described in Section 6.2. The only difference is that now the support of the marginal valuation distributions is the interval between the lowest and the highest transaction prices observed in the data. In light of the findings in Section 6, I set $\kappa_n = n/\ln n$, $I_n = \{(0, n/\ln n), (\sqrt{1/\ln n}, 0), (\sqrt{1/2\ln n}, n/2\ln n)\}$, and $\eta = 0$. I also set $\mathbf{B} = 5$ because in all simulations in Section 6, the value of $\mathbf{B}$ does not seem to change results much, and a larger $\mathbf{B}$ typically leads to longer computing time. As a comparison, I also consider a version of my test without imposing FOSD.

The results are shown in Figure 4. The dashed lines with circle markers are the upper and lower bounds of the pointwise 90% confidence set with FOSD, while the ones with cross markers are bounds of the confidence set without FOSD. The confidence set with FOSD is much narrower compared to the one without FOSD. I also compute the estimated valuation distribution assuming symmetry and independence, which is shown by the solid line. Imposing FOSD allows me to exclude the solid line, at least at certain points, from the conference set. Without FOSD, I cannot exclude any part of the solid line. This shows that imposing the shape restriction (FOSD) can greatly improve inference.

8. Conclusion

This paper proposes a test of a shape restriction on an infinite-dimensional parameter that is identified by unconditional moment equalities. This test can be inverted to construct confidence sets. It is very general and has many applications. One attractive feature of the test is that it exploits convexity to strike a balance between computing time and the need to approximate the local parameter space. Specifically, if partial identification is a concern, my method involves optimization problems that have a much lower dimension compared with the method proposed in CNS. This can be crucial if the dimension of the sieve space is high and/or in structural estimation where moments are difficult to evaluate. I illustrate my method in an empirical application to ascending auctions. The results show that imposing shape restrictions can greatly improve inference.

APPENDIX A: NOTATION LIST

The parameter space Θ is equipped with a norm $\|\cdot\|_s$ and d_s is the metric induced by $\|\cdot\|_s$. Let $\|\cdot\|_F$ be a chosen norm on the parameter space Θ under which the convergence property of $\hat{\Theta}_n(R)$ is important. Denote the metric induced by $\|\cdot\|_F$ as d_F. Notice that $\|\cdot\|_F$ can depend on F. For example, if $\theta = (\theta^P, \theta^N)$ where θ^P is a finite-dimensional real vector and θ^N is a vector-valued function on $\mathcal{W}$, then a common choice is $\|\theta\|_F^2 = \|\theta^P\|_E^2 + \|\theta^N\|_{2,F}^2$, where $\|\cdot\|_E$ is the Euclidean norm and $\|\theta^N\|_{2,F}^2 = \int \|\theta^N(W)\|_E^2 \, dF$ is the usual L^2 norm under the distribution F. Also define

$$\Theta_{n,F}(R) = \arg\min_{\theta \in \Theta_n \cap R} Q_F(\theta)$$

to be the set of minimizers of the population criterion function on the sieve space. For any function f on Θ,

$$\arg\inf_{\theta \in \Theta_n \cap R} f(\theta)$$

$$= \Big\{ \theta \in \Theta : \exists \{\theta_k\}_{k=1}^\infty \subset \Theta_n \cap R \text{ such that } \lim_{k\to\infty} f(\theta_k) = \inf_{\theta \in \Theta_n \cap R} f(\theta), \lim_{k\to\infty} d_s(\theta, \theta_k) = 0 \Big\}.$$

It is nonempty and lies in $\Theta_n \cap R$ if $\Theta_n \cap R$ is compact under d_s. If, in addition, f is continuous, then $\arg\inf_{\theta \in \Theta_n \cap R} f(\theta) = \arg\min_{\theta \in \Theta_n \cap R} f(\theta)$.

Define the first-order directional derivative of $\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \cdot)$ at θ in the direction Δ as

$$\frac{d\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)}{d\theta}[\Delta] = \lim_{\tau \downarrow 0} \frac{\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta + \tau\Delta) - \mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)}{\tau}.$$

Let $\ell^\infty(\Theta \times \mathbf{T})$ be the space of bounded functions on $\Theta \times \mathbf{T}$ equipped with the standard sup norm $\|\cdot\|_\infty$. Let d_∞ be the metric induced by $\|\cdot\|_\infty$. $\Gamma_{\kappa,F,m,R} : \ell^\infty(\Theta \times \mathbf{T}) \to \mathbb{R}$ is a function indexed by $\kappa \geq 0$, $F \in \mathcal{F}$, $m \in \mathbb{Z}^+$, and $R \in \mathcal{R}$ such that

$$\Gamma_{\kappa,F,m,R}(\omega) = \inf_{\tilde{\theta} \in \Theta_{m,F}(R)} \Big\{ \inf_{\theta \in B_F^\kappa(\tilde{\theta}) \cap \Theta_m \cap R} \int_{\mathbf{T}} \Big[\omega(\tilde{\theta}, \mathbf{t}) + \frac{\sqrt{m} d\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \tilde{\theta})}{d\theta}[\theta - \tilde{\theta}] \Big]^2 \, d\mu(\mathbf{t}) \Big\}.$$

Notice that if $\kappa_1 > \kappa_2$, then $\Gamma_{\kappa_1,F,m,R}(\omega) \leq \Gamma_{\kappa_2,F,m,R}(\omega)$ because inf in the curly bracket is taken over a larger set under κ_1. The theory of this paper involves showing that the test statistic and the bootstrap statistic can be approximated asymptotically by $\Gamma_{\kappa,F,m,R}(\cdot)$ applied to the empirical process and the bootstrap empirical process, respectively. Lastly, denote the weak limit of $\mathbb{G}_{n,F}(\theta, \mathbf{t})$ by $\mathbb{G}_F$ and Table 2 contains other important notation used in this appendix.

APPENDIX B: ASSUMPTIONS FOR UNIFORM SIZE CONTROL

ASSUMPTION B.1. (i) *For every $R \in \mathcal{R}$, $\Theta \cap R$ is compact under d_s*; (ii) *there exist $C_1, C_2 \in (0, \infty)$ such that $C_1 d_{w,F} \leq d_F \leq C_2 d_s$ for every $F \in \mathcal{F}$.*

TABLE 2. Some important notation.

$B_F^\epsilon(\theta)$	$\{\tilde\theta \in \Theta : d_F(\tilde\theta, \theta) \leq \epsilon\}$
$d_{w,F}(a, b)$	$\sqrt{\int_{\mathbf{T}}[\mathbb{E}_F\rho_{\mathbf{t}}(W_i, a) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i, b)]^2\, d\mu(\mathbf{t})}$ for $a, b \in \Theta$
$d_F(a, B)$	$\inf_{b\in B} d_F(a, b)$ for $a \in \Theta$ and $B \subset \Theta$
$d_s(a, B)$	$\inf_{b\in B} d_s(a, b)$ for $a \in \Theta$ and $B \subset \Theta$
$d_{w,F}(a, B)$	$\inf_{b\in B} d_{w,F}(a, b)$ for $a \in \Theta$ and $B \subset \Theta$
$\Theta_{n,F}^\epsilon(R)$	$\{\theta \in \Theta_n \cap R : d_F(\theta, \Theta_{n,F}(R)) \leq \epsilon\}$
Θ_F^ϵ	$\{\theta \in \Theta : d_F(\theta, \Theta_F) \leq \epsilon\}$
$a \vee b$	Maximum of a and b
$\downarrow$	Converge from above
$\rightarrow$	Converge to

ASSUMPTION B.2. *The sieve space Θ_n satisfies the following conditions.* (i) *For every n, $\Theta_n \subset \Theta$ is finite-dimensional, convex, and closed under d_s.* (ii) *As $n \to \infty$, $\sup_{R\in\mathcal{R}} \sup_{\theta\in\Theta\cap R} d_s(\theta, \Theta_n \cap R) = o(1)$.* (iii) *As $n \to \infty$, $\sup_{R\in\mathcal{R}} \sup_{\theta\in\Theta\cap R} d_s(\theta, \Theta_n \cap R) = o(n^{-1/2})$.*

ASSUMPTION B.3. (i) *For any $\epsilon > 0$, the following quantity,*

$$S_n(\epsilon) = \sup_{(F,R)\in\mathcal{J}}\left[\inf_{\theta\in\Theta_n\cap R:d_F(\theta,\Theta_{n,F}(R))>\epsilon} Q_F(\theta) - \inf_{\theta\in\Theta_n\cap R} Q_F(\theta)\right]$$

satisfies that $\sqrt{n}S_n(\epsilon) \to \infty$ if $n \to \infty$. (ii) *There exist $\epsilon > 0$ and $\nu \geq 1$ such that*

$$\zeta_n = \sup_{(F,R)\in\mathcal{J}} \sup_{\theta\in\Theta_{n,F}^\epsilon(R):d_{w,F}(\theta,\Theta_{n,F}(R))\neq 0} \frac{d_F\big(\theta, \Theta_{n,F}(R)\big)^\nu}{d_{w,F}\big(\theta, \Theta_{n,F}(R)\big)}$$

is finite for every $n > 0$.

ASSUMPTION B.4. (i) *There exists $\epsilon > 0$ such that $\frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i,\theta_1)}{d\theta}[\theta_2 - \theta_1]$ exists for every $\theta_1, \theta_2 \in \Theta_F^\epsilon$, $\mathbf{t} \in \mathbf{T}$, and $F \in \mathcal{F}$.* (ii) *There exist $\epsilon > 0$ and $\xi_n > 0$ such that $\xi_n \to 0$, $\sqrt{n}\xi_n \to \infty$ as $n \to \infty$ and*

$$\sup_{(F,R)\in\mathcal{J}} \sup_{\theta_1,\theta_2\in\Theta_{n,F}^\epsilon(R):d_F(\theta_1,\theta_2)\leq\xi_n} \sup_{\mathbf{t}\in\mathbf{T}}\left|\mathbb{E}_F\rho_{\mathbf{t}}(W_i, \theta_2) - \mathbb{E}_F\rho_{\mathbf{t}}(W_i, \theta_1) - \frac{d\mathbb{E}_F\rho_{\mathbf{t}}(W_i, \theta_1)}{d\theta}[\theta_2 - \theta_1]\right|$$

$$= o\big(n^{-1/2}\big).$$

(iii) *There exists $C_D < \infty$ such that for all $F \in \mathcal{F}$*

$$\sup_{\mathbf{t}\in\mathbf{T}} \mathbb{E}_F\big[\rho_{\mathbf{t}}(W_i, \theta_1) - \rho_{\mathbf{t}}(W_i, \theta_2)\big]^2 \leq C_D d_F(\theta_1, \theta_2)^2.$$

ASSUMPTION B.5. *The tuning parameter κ_n satisfies $\kappa_n \to \infty$, $\kappa_n \ln\ln n/n \to 0$, and $\zeta_n/\sqrt{\kappa_n} = o(\xi_n^\nu)$.*

ASSUMPTION B.6. (i) $\Theta \cap S$ is convex and compact under d_s. (ii) ϕ is linear and $\phi(\theta_F)$ lives in a known compact set $\mathbb{K} \subset \mathbb{R}^{D_\phi}$, for all $F \in \mathcal{F}$. (iii) $\sup_{a \in \mathbb{K}} \sup_{\theta \in \Theta \cap R_{a,\phi,S}} d_s(\theta, \Theta_n \cap R_{a,\phi,S}) = o(n^{-1/2})$.

ASSUMPTION B.7. Let $\Gamma_{m,J} = \Gamma_{\xi_m,F,m,R}$ where $J = (F, R)$ and let $C_F(1 - \alpha, \Gamma_{m,J})$ be the $1 - \alpha$th quantile of $\Gamma_{m,J}(\mathbb{G}_F)$ and

$$\mathcal{J}_n^{\epsilon}(\alpha) = \{J \in \mathcal{J} : C_F(1 - \alpha, \Gamma_{m,J}) > \epsilon \text{ if } m \geq n\}.$$

(i) For every $\epsilon > 0$, if $\eta \downarrow 0$, then

$$\limsup_{n \to \infty} \sup_{J \in \mathcal{J}_n^{\epsilon}(\alpha)} \mathbb{P}_F\big(C_F(1 - \alpha - \eta, \Gamma_{n,J}) - \eta < \Gamma_{n,J}(\mathbb{G}_F) \leq C_F(1 - \alpha + \eta, \Gamma_{n,J}) + \eta\big) \to 0.$$

(ii) For every $\epsilon_n > 0$ and $\epsilon_n \to 0$, $\limsup_{n \to \infty} \sup_{(F,R) \in \mathcal{J} \setminus \mathcal{J}_n^{\epsilon_n}(\alpha)} \mathbb{P}_F(T_n(R) > 0) < \alpha$.

Assumption B.1(i) holds after regularizing the parameter space following Santos (2012). As in Santos (2012), the choice of Θ depends on R. For example, if R involves derivatives of θ, Θ needs to be chosen accordingly. For more details, see Section S.2 in the Online Supplemental Material. Assumption B.1(ii) is satisfied with many commonly used norms. For example, if d_s is the sup metric and d_F is induced by the L^2 norm under F, then standard smoothness assumptions on $\rho_{\mathbf{t}}(W_i, \theta)$ imply Assumption B.1(ii). Assumption B.1 ensures that $Q_F(\cdot)$ is continuous under d_s and that $\inf_{\theta \in \Theta \cap R} Q_F(\theta) = \min_{\theta \in \Theta \cap R} Q_F(\theta)$. It implies that $\min_{\theta \in \Theta \cap R} Q_F(\theta) > 0$ under any fixed alternative, which is crucial for the power result of the test. Assumption B.2(i) is a mild requirement. Together with Assumption B.1(i), it implies the compactness of $\Theta_n \cap R$ under d_s.[12] By the definition of $\arg\inf$, this further implies that $\hat{\Theta}_n(R)$ is nonempty and lies in $\Theta_n \cap R$. Convexity of $\Theta_n \cap R$ can be relaxed if $\frac{d\mathbb{E}_F \rho_{\mathbf{t}}(W_i, \theta)}{d\theta}[\Delta]$ is linear in Δ and uniformly bounded. Assumptions B.2(ii) and (iii) require the sieve spaces to approximate $\Theta \cap R$ well enough under d_s uniformly in $R \in \mathcal{R}$. Under Assumption B.1(ii), this means that the sequence of sieve spaces approximates $\Theta \cap R$ well under d_F uniformly in $(F, R) \in \mathcal{J}$. If d_s is the sup metric and one regularizes Θ following Newey and Powell (2003) and Santos (2012), then the sieve spaces can be spanned by orthogonal polynomials or splines with sufficiently many orders or knots. See Newey (1997) for more discussion. Notice, unlike in Hong (2017) and Santos (2012), there is no restriction on how the boundary points of Θ can be approximated by interior points.

Assumption B.3(i) is similar to Assumption 4.1(ii) in CNS. It allows that as n increases, $\Theta_{n,F}(R)$ becomes less well separable, that is, $S_n(\epsilon)$ can converge to 0 for a given ϵ. However, the convergence rate must be slower than $\sqrt{n}$. This assumption rules out weak identification. It, however, does not require $\sup_{\theta \in \Theta_{n,F}(R)} d_F(\theta, \Theta_F \cap R) \to 0$ as $n \to \infty$, that is, the minimizers of $Q_F(\theta)$ on the sieve space are allowed to be far away from the true identified set under d_F.

Assumption B.3(ii) is otherwise a standard assumption in the sieve literature except that I also take sup across all $F \in \mathcal{F}$. Notice that $d_{w,F}$ is related to the weak

---

[12] If one is only concerned with size control, one can directly assume that $\Theta_n \cap R$ is compact for every n and $R \in \mathcal{R}$. Then Assumption B.1(i) is not needed.

norm considered in Ai and Chen (2003). There are several differences. First, since I am focusing on models with a large number of unconditional moments, $d_{w,F}$ is based on the integration of these unconditional moments under μ. Second, $d_{w,F}$ does not allow for a general weight matrix, but it can be extended in this direction. Third, $d_{w,F}$ does not involve the directional derivatives at the true value of the unknown parameter. This is more convenient for models with partial identification. It is also worth pointing out that in nonparametric IV models, one can choose weight functions and μ such that $d_{w,F}$ is arbitrarily close to the weak norm defined in Ai and Chen (2003). Also, ζ_n is closely related to the measure considered in CNS, who transform conditional moments into countably infinitely many moments using basis functions. In fact, given the set of basis functions, one can construct w and μ such that my method essentially uses the same set of unconditional moments as in CNS. Then ζ_n with $\nu = 1$ is different from its counterpart in CNS under $r = 2$ in two ways. First, it uses all the moments for any finite n, while CNS use a finite subset of these moments. Second, within the finite subset of moments used in CNS, ζ_n typically give each moment a different weight compared to CNS.

If $\mathcal{F}$ is known, ζ_n can be directly calculated. Alternatively, if $\mathcal{F} = \{F_0\}$ and ν is known, one may estimate ζ_n by replacing d_F and $d_{w,F}$ by their sample analog and replacing $\Theta_{n,F}(R)$ by a consistent estimator. As a special case, if $\nu = 1$, d_F is the L^2 norm, $\rho_{\mathbf{t}}$ is linear in θ for all $\mathbf{t}$, and Θ_n is a linear space with a finite number of basis functions, then one can estimate ζ_n following the method proposed in Chen and Christensen (2015) if θ_F is point-identified.

Assumptions B.4(i) and (ii) imply that the moment conditions can be approximated by their linear expansion in a neighborhood of $\Theta_{n,F}(R)$. Combined with other assumptions, they rule out cases where all directional derivatives are 0 at the boundary of $\Theta_{n,F}(R)$, that is, local identification failure as considered in Dovonon and Renault (2013) and Lee and Liao (2018). Assumption B.4(iii) accompanied by Definition 5.1(iii) implies that the empirical process $\mathbb{G}_{n,F}(\theta, \mathbf{t})$ is asymptotically equicontinuous in θ.

Assumptions B.1–B.4 are sufficient to guarantee that the test statistic is approximated asymptotically by $\Gamma_{\xi_n,F,n,R}(\mathbb{G}_{n,F})$ under the null hypothesis. Assumption B.5 is needed to validate the bootstrap statistic. It guarantees that the rescaled moments can be approximated by their first-order expansion around the identified set. If the model is linear in θ, ξ_n can be any $o(1)$ sequence and κ_n can be any sequence that does not diverge too fast. If the model is nonlinear and $\nu = 1$, usually ξ_n can be any $o(n^{-1/4})$ sequence. Then κ_n needs to satisfy $\zeta_n/\sqrt{\kappa_n} = o(n^{-1/4})$ and $\kappa_n \ln\ln n/n \to 0$. In some cases, such κ_n may not exist. In particular, these requirements rule out severely ill-posed nonlinear models. Similar assumptions can also be found in the semiparametric conditional moment literature if endogenous variables enter the unknown functions. For an example, see Chen and Pouzo (2009). It is worth noting that if $I_n = \{(0, \kappa_n)\}$, this assumption can be relaxed to only require $\zeta_n/\sqrt{\kappa_n} = o(1)$. Assumption B.6 guarantees that the confidence set for $\phi(\theta_F)$ is valid. It is a version of Assumption B.2 with $\mathcal{R} = \{R_{a,\phi,S} : a \in \mathbb{K}\}$.

Assumption B.7 is required only if one would like to set $\eta = 0$. Assumption B.7(i) holds if there exists $\delta > 0$ such that uniformly in $J \in \mathcal{J}$ and sufficiently large m, $\Gamma_{m,J}(\mathbb{G}_F)$ has a density bounded away from 0 and from above on $((C_F(1 - \alpha, \Gamma_{m,J}) - \delta) \vee 0, C_F(1 - \alpha, \Gamma_{m,J}) + \delta)$. Assumption B.7(ii) handles the case where $C_F(1 - \alpha, \Gamma_{m,J})$ is equal to or arbitrarily close to 0. It is related to Assumption A.2 in Bugni, Canay, and Shi (2017).

References

Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71 (6), 1795–1843. [611, 633]

Andrews, D. W. (2012), "Similar-on-the-boundary tests for moment inequalities exist, but have poor power." Cowles Foundation Discussion Paper 1815R. Available at https://ssrn.com/abstract=2016447. [617]

Andrews, D. W. K. and X. Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81 (2), 609–666. [610]

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78 (1), 119–157. [610]

Aradillas-López, A., A. Gandhi, and D. Quint (2013), "Identification and inference in ascending auctions with correlated private values." *Econometrica*, 81 (2), 489–534. [614, 615, 628, 629]

Aradillas-López, A., A. Gandhi, and D. Quint (2016), "A simple test for moment inequality models with an application to English auctions." *Journal of Econometrics*, 194 (1), 96–115. [628]

Athey, S., J. Levin, and E. Seira (2011), "Comparing open and sealed bid auctions: Evidence from timber auctions." *Quarterly Journal of Economics*, 126 (1), 207–257. [628]

Bierens, H. J. (1990), "A consistent conditional moment test of functional form." *Econometrica*, 58 (6), 1443–1458. [613]

Bierens, H. J. and H. Song (2012), "Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method." *Journal of Econometrics*, 168 (1), 108–119. [611, 612, 614]

Blundell, R., X. Chen, and D. Kristensen (2007), "Semi-nonparametric IV estimation of shape-invariant Engel curves." *Econometrica*, 75 (6), 1613–1669. [610, 611, 613]

Bugni, F. A., I. A. Canay, and X. Shi (2017), "Inference for subvectors and other functions of partially identified parameters in moment inequality models." *Quantitative Economics*, 8 (1), 1–38. [633]

Canay, I. A., A. Santos, and A. M. Shaikh (2013), "On the testability of identification in some nonparametric models with endogeneity." *Econometrica*, 81 (6), 2535–2559. [610, 612]

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2014), "Local identification of nonparametric and semiparametric models." *Econometrica*, 82 (2), 785–809. [611]

Chen, X. and T. Christensen (2015), "Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation." Cowles Foundation discussion paper. [633]

Chen, X. and D. Pouzo (2009), "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals." *Journal of Econometrics*, 152 (1), 46–60. [611, 633]

Chen, X. and D. Pouzo (2012), "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals." *Econometrica*, 80 (1), 277–321. [611]

Chen, X. and D. Pouzo (2015), "Sieve Wald and QLR inferences on semi/nonparametric conditional moment models." *Econometrica*, 83 (3), 1013–1079. [611]

Chernozhukov, V., W. K. Newey, and A. Santos (2015), "Constrained conditional moment restriction models." Cemmap Working Paper CWP59/15. Available at arXiv:1509.06311. [611]

Chetverikov, D. (2019), "Testing regression monotonicity in econometric models." *Econometric Theory*, 35 (4), 729–776. [611]

Chetverikov, D. and D. Wilhelm (2017), "Nonparametric instrumental variable estimation under monotonicity." *Econometrica*, 85 (4), 1303–1320. [611, 613]

Coey, D., B. Larsen, K. Sweeney, and C. Waisman (2017), "Ascending auctions with bidder asymmetries." *Quantitative Economics*, 8 (1), 181–200. [614, 615]

Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011), "Nonparametric instrumental regression." *Econometrica*, 79 (5), 1541–1565. [611]

Dovonon, P. and E. Renault (2013), "Testing for common conditionally heteroskedastic factors." *Econometrica*, 81 (6), 2561–2586. [633]

Fan, Y., M. He, and T. Li (2015), "A sensitivity analysis in the affiliated private value auction model with incomplete set of bids." Unpublished manuscript, University of Washington. Available at https://editorialexpress.com/cgi-bin/conference/download. cgi?db_name=CESG2015&paper_id=59. [627]

Fang, Z. and J. Seo (2019), "A general framework for inference on shape restrictions." Unpublished manuscript, Texas A&M University. Available at arXiv:1910.07689. [611]

Freyberger, J. and J. L. Horowitz (2015), "Identification and shape restrictions in nonparametric instrumental variables estimation." *Journal of Econometrics*, 189 (1), 41–53. [611]

Freyberger, J. and B. Reeves (2018), "Inference under shape restrictions." Unpublished manuscript, University of Wisconsin-Madison. Available at https://www.ssc.wisc.edu/ ~jfreyberger/Shape_Inference_Freyberger_Reeves.pdf. [611]

Gandhi, A., Z. Lu, and X. Shi (2013), "Estimating demand for differentiated products with error in market shares: A moment inequalities approach." Unpublished manuscript, University of Wisconsin-Madison. Available at https://www.ssc.wisc.edu/ ~agandhi/homepage/Amit_Gandhi_files/gandhi.lu.shi_01-31-13.pdf. [610, 617, 625]

Gutknecht, D. (2016), "Testing for monotonicity under endogeneity: An application to the reservation wage function." *Journal of Econometrics*, 190 (1), 100–114. [610, 611, 613]

Hall, P. and J. Horowitz (2005), "Nonparametric methods for inference in the presence of instrumental variables." *The Annals of Statistics*, 33 (6), 2904–2929. [611]

Hong, S. (2017), "Inference in semiparametric conditional moment models with partial identification." *Journal of Econometrics*, 196 (1), 156–179. [611, 612, 616, 617, 625, 632]

Horowitz, J. L. and S. Lee (2017), "Nonparametric estimation and inference under shape restrictions." *Journal of Econometrics*, 201 (1), 108–126. [611]

Komarova, T. (2013), "Partial identification in asymmetric auctions in the absence of independence." *The Econometrics Journal*, 16 (1), S60–S92. [615]

Lee, J. H. and Z. Liao (2018), "On standard inference for GMM with local identification failure of known forms." *Econometric Theory*, 34 (4), 790–814. [633]

Linton, O., K. Song, and Y.-J. Whang (2010), "An improved bootstrap test of stochastic dominance." *Journal of Econometrics*, 154 (2), 186–202. [621]

Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79 (1), 147–168. [632]

Newey, W. K., J. L. Powell, and F. Vella (1999), "Nonparametric estimation of triangular simultaneous equations models." *Econometrica*, 67 (3), 565–603. [611, 624]

Newey, W. K. and J. L. Powell (2003), "Instrumental variable estimation of nonparametric models." *Econometrica*, 71 (5), 1565–1578. [611, 621, 632]

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015), "Moment inequalities and their application." *Econometrica*, 83 (1), 315–334. [610]

Santos, A. (2012), "Inference in nonparametric instrumental variables with partial identification." *Econometrica*, 80 (1), 213–275. [611, 612, 616, 617, 621, 623, 624, 632]

Stinchcombe, M. B. and H. White (1998), "Consistent specification testing with nuisance parameters present only under the alternative." *Econometric Theory*, 14 (3), 295–325. [613]

Tao, J. (2014), "Inference for point and partially identified semi-nonparametric conditional moment models." Unpublished manuscript, University of Wisconsin-Madison. Available at https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=ESWC2015&paper_id=2286. [611]

Zhu, Y. (2020), "Supplement to 'Inference in nonparametric/semiparametric moment equality models with shape restrictions'." *Quantitative Economics Supplemental Material*, 11, https://doi.org/10.3982/QE1023. [612]

Zhu, Y. and S. Grundl (2014), "Nonparametric tests in moment equality models with an application to infer risk aversion in first-price auctions." Unpublished manuscript. Available at https://drive.google.com/file/d/0B5Rw7pzGrkeXWnFYalFEcnFMUU0/view. [610, 615]